iScience

Vocal Emotion Recognition: A Comparison of Singers and Instrumentalists, Amateurs and Professionals

--Manuscript Draft--

Manuscript Number:	ISCIENCE-D-25-11893R1
Full Title:	Vocal Emotion Recognition: A Comparison of Singers and Instrumentalists, Amateurs and Professionals
Article Type:	Research Article
Corresponding Author:	Christine Nussbaum Friedrich Schiller University Jena Jena, GERMANY
First Author:	Christine Nussbaum
Order of Authors:	Christine Nussbaum
	Jessica Dethloff
	Annett Schirmer
	Stefan R. Schweinberger
Abstract:	Musicians outperform non-musicians in vocal emotion recognition, presumably due to differences in auditory sensitivity. However, the current literature is inconclusive regarding differential effects of specific types of musical activity. Given the tight link between expression and perception in vocal emotional communication, we assessed possible effects of whether music is expressed vocally or not, and whether music is performed on an amateur or professional level. Importantly, however, we predicted that vocal emotion recognition would be unaffected by the type and amount of musical activity because current evidence argues against a causal role of formal musical education. We compared emotion recognition performance of singers (N= 45) vs. instrumentalists (N = 43) and professional musicians (N = 40) vs. amateurs (N = 88) vs. non-musicians (N = 38), based on short vocal utterances expressing happiness, pleasure, fear, or sadness. Using both frequentist and Bayesian inference, we found the predicted nulleffects for singers vs. instrumentalists, and professionals vs. amateurs. Evidence for an advantage in amateurs vs. non-musicians was inconclusive. Across groups, we replicated the consistent link between vocal emotion recognition and auditory sensitivity. Overall, the current work aligns with the perspective that musicians' advantage in recognizing vocal emotions is rooted in auditory sensitivity, rather than specific types of musical activities or formal training.
Additional Information:	
Question	Response
Standardized datasets A list of datatypes considered standardized under Cell Press policy is available here. Does this manuscript report new standardized datasets?	No
Original code Does this manuscript report original computer code, algorithms, or computational models?	No

FRIEDRICH-SCHILLER-UNIVERSITÄT JENA
Institut für Psychologie
Lehrstuhl für Allgemeine Psychologie und kognitive Neurowissenschaften

Universität Jena · Lehrstuhl für Allgemeine Psychologie · Institut für Psychologie · 07737 Jena

Univ.-Prof. Dr. Stefan R. Schweinberger
Christine Nussbaum, PhD

Am Steiger 3/Haus 1
07743 Jena

Telefon: 0 36 41 9-45181
Telefax: 0 36 41 9-45182
E-Mail: stefan.schweinberger@uni-jena.de
christine.nussbaum@uni-jena.de

To the editors of
iScience

Jena, 24. September 2025

Re: Resubmission of Manuscript ISCIENCE-D-25-11893 ; your letter per e-mail dated August 29, 2025

Dear Dr. Rituparna Chakrabarti, Deputy Editor of *iScience*,

thank you for your letter regarding the above manuscript, and for inviting us to submit a revision of our work. Attached please find the revised manuscript by Christine Nussbaum, Jessica Dethloff, Annett Schirmer, and Stefan R. Schweinberger entitled “Vocal Emotion Recognition: A Comparison of Singers and Instrumentalists, Amateurs and Professionals” for possible publication as a research article in *iScience*.

We included a response letter which gives details on how we addressed each individual suggestion. In short, we considered the full set of very constructive and helpful suggestions by yourself and the reviewer and implemented the corresponding revisions in the manuscript and the supplemental materials. We are very grateful to you and the reviewers for the detailed feedback.

Please note that in reviewer point 7, there seemed to be confusion regarding the publication date of the preregistration. We are very keen to resolve this issue and assure full transparency from our side on this matter.

In line with the journal guidelines, we also prepared a Graphical Abstract, the Highlights and a document with the STAR methods to the best of our knowledge. The Graphical Abstract was designed using Canva Pro Version (https://www.canva.com/).

In our opinion, the implementation of suggestions raised by the reviewers improved our paper and made it more comprehensible. We hope you agree that the manuscript benefited from revision and is now in good shape for inclusion in *iScience*.

Yours sincerely,
Christine Nussbaum, Jessica Dethloff, Annett Schirmer and Stefan R. Schweinberger

Response to Reviews:

1. **Reviewer #1: The authors investigate vocal emotion recognition (VER) performance in musicians with different levels of training and activity in different domains (instrumentalists and singers). They presumed to find no influences of either factor, which was confirmed. However, they found that VER correlated with measures of auditory sensitivity. They conclude that VER operates at a perceptual level (auditory sensitivity) and cognitive levels (musicality, musical activity) may play little role in this respect. The findings do contribute to the literature. Whether they really close a gap, as the authors claim, remains to be seen. I think it is a bold statement to say that a gap has been closed. It feels a bit overambitious to me. Nevertheless, an interdisciplinary community could benefit from learning what musicality can mean and what not.**
   Reviewer #2: The authors find that vocal emotion recognition does not vary between singers and instrumental musicians, or between amateurs and professionals. The finding is important because if music training is the causal factor in links between musicianship and vocal emotion recognition, one would expect differences in performance based on the type of musical background, particularly whether it focuses on the voice, and the extent of training/performing.
   Reviewer #3: Thank you for the opportunity to review this interesting manuscript. Overall, I find it to be a well-written and well-justified study. The design builds on a previously published experiment and aims to test specific hypotheses regarding vocal emotion recognition across different musical subgroups. The study was preregistered with nine hypotheses that are tested for, and the target sample size (N=40 per group) was met or slightly exceeded. Recruitment procedures, inclusion/exclusion criteria, and compensation protocols were implemented as described. In addition, exploratory analyses were conducted, which is acceptable as long as they are clearly acknowledged as such. The study aligns with previous work, that musicians expertise in vocal emotion recognition is explained by auditory sensitivity, rather than musical training. The study extends a previous study comparing professionals and non-musicians, by including a group of amateurs. Comparisons are also made between singers and instrumentalists.

Response: We thank all three reviewers for their effort and their careful evaluation of this manuscript. Below, we will respond to each point in detail.

2. **Reviewer #1: Humans operate at different, but highly intertwined levels, perceptual, cognitive, social. There is constant interaction between those. The authors conducted a study to suggest that VER is likely "programmed" to some degree in the nervous system. If it could be modified by learning, music is less a candidate than it might seem. This is a fair conclusion from the literature, to which the current study attempts to contribute. Nevertheless, participants respond in full consciousness, thus using their cognitive abilities to allow for inferences about their perceptual**

skills. So, in other words, there is no direct, albeit converging evidence to sustain the hypothesis of enhanced auditory sensitivity per se to explain individual differences in VER, irrespective of musical training and modality of training (singing vs. playing instruments). Please explain, why musicianship should play a role in such short utterances. I am aware that the "springboard" seems to be initial positive findings, but does that warrant continued investigations? Why should vocalists have an advantage? But the test amounts to show they do not have one. Why it is important to confirm this? Any indication in the literature expect the opposite? In other words, is this an original hypothesis or an exploration?

Response: You raise several valid questions here. First, we fully agree that emotional inferences like the ones we target here are the results of intertwined perceptual, cognitive and social processes, which all contribute to vocal perception performance. Based on behavioral data alone, they can hardly be disentangled. We do find a consistent link between auditory sensitivity and individual differences in VER, but we never claimed that this is a purely perceptual process. Quite the contrary, based on previous insights from an EEG-study (Nussbaum et al., 2023), we consider it very plausible that cognitive abilities, especially how auditory information is used to make a conscious decision, plays a role. We mention this work in the discussion for the interested reader. Importantly, this pattern does not seem to be affected by the type or amount of musical training, which was our focus in the present work.

Regarding your question why musicianship should play a role in such short utterances: the performance difference between musicians and non-musicians in vocal emotion perception based on very short utterances has been shown repeatedly and is summarized in two reviews (Nussbaum et al 2021, Martins et al 2021) which we mention in the beginning of our introduction.

Regarding your questions why we study this with vocalists and instrumentalists, we have now added/refined two paragraphs in the introduction. Details can be found in our response to point 4. And indeed, this is an original (and preregistered) hypothesis, based on a careful evaluation of the previous literature.

3. **The psychoacoustic transformation of stimuli materials is an interesting feature of the study. It suggests that no sensitivity can compensate for a reduction of stimulus information that leaves the participants merely with timbre differences. However, some emotions seem to be affected more or less strongly by morph type (interaction), which seems to make the picture more complicated. Indeed, the selection of emotions (p.10) appears not intuitive. The rationale to use happiness and pleasure (alongside fear and sadness) as distinct categories should be explained. Is there a circumplex-idea in the background? Could it be confirmed that the four emotions occupy four quadrants? Otherwise, there could be a confound arising from**

similarity and differences of underlying dimensions of the four emotions. - Please explain.

Response: Yes, indeed, a reduction of stimulus information to timbre information does have an impact on the recognition of all four emotions (although to different degrees) across all participants. However, participants are able to use timbre information to some degree, since emotion recognition is still above chance in this condition in all emotions (cf. Figure 3). And indeed, emotions are affected to different degrees by this manipulation, as indicated by the Morph Type x Emotion interaction. We have found this interaction consistently across a number of studies, also with partly different stimuli (e.g. Nussbaum 2022, SCAN; Nussbaum 2023, Brain Sciences, and Nussbaum 2024, BJOP). In short, we always see a similar pattern: in general, pitch (F0) information plays a more dominant role, relative to timbre. However, in emotions with high intensity (especially happiness), the difference is much larger. In emotions with lesser intensity (like pleasure and sadness) the contribution of timbre and F0 information is more balanced. We have not discussed this in full detail here, because it was not the focus of the present study, but you find a detailed discussion here: https://doi.org/10.1093/scan/nsac033. To make this transparent to the readers now, we added the following: *"Note that the effects of the F0 and timbre manipulation on emotion recognition also provided a complete replication of the pattern observed in professional musicians and non-musicians in our previous study (Nussbaum et al., 2024). For an even more detailed reflection on the roles of F0 and timbre for emotion recognition, both on a behavioral and neural level (irrespective of musicality), please also refer to Nussbaum et al. (2022)." (page 28/29)*

Concerning the choice of emotions: Indeed, the rationale behind the choice was to balance positive and negative emotions and their emotional intensity. When using basic emotions only, happiness as the only positive emotion is usually the "oddball", which can be problematic. Thus, we added a second one, which was pleasure. The stimulus material was validated in a previous rating study to ensure that the four emotions occupy the four quadrants of valence/intensity. We added this information in the methods section: "*A prior validation study with 20 raters confirmed that the two positive and two negative emotions had different degrees of emotional intensity (happiness > pleasure, $t(19)=9.57$, $p < 0.001$ and fear > sadness, $t(19)=6.58$, $p < 0.001$).* " (page 10)

4. **I wonder whether VER could be "boiled down" to the ability to recognize and differentiate pitch contours vocal utterances, with anything going up indicating arousal, and downward indicating relaxation to guess the correct emotion on the basis of contour alone seems rather plausible. So, in our daily interactions, we use this skill myriads of times. Why should even 10.000 hours of music training in 10 years (to reach expertise) change anything in this basic skill? So, why are the findings not trivial after all? Please explain, especially in the light of testing a Null hypothesis.**

Response: Our findings displayed in Figure 3 show that emotional inferences are possible based on timbre information alone (although to a lesser degree than based on pitch information). Further, emotion perception performance is best, when both pitch + timbre information is available (in the Full condition). Thus, while pitch contour undoubtedly plays a dominant role, it is not the only acoustic parameter that signals emotional information and can be picked up by listeners. Further, while we show that pitch contour plays a big role in general, a conclusion like “upward = higher arousal” and “downward = less arousal” seems oversimplified. It could also be something like “high pitch variation = higher arousal” and “low pitch variation = lower arousal”. Please note that the specific way in which emotion is inferred based on pitch cues was not the focus of the present study, so we refrain from any speculation about this matter. We hold, however, the viewpoint that VER cannot simply be boiled down to the perception of pitch contour direction alone.

Regarding your question why even 10.000 hours of music training in 10 years would make a difference: Musicians have a robust advantage over non-musicians in vocal emotion recognition and Juslin and Laukka (2003) showed in an extensive meta-analysis that emotions expressed by voices and by music share a similar acoustic code. Based on this, early theoretical frameworks like the OPERA hypothesis (Patel 2011) proposed a causal effect of musical training on non-musical skills, if five specific conditions are met: Overlap (between the neural networks involved), (2) Precision (i.e. high auditory-motor demands), (3) Emotion (i.e. musical activity has to be considered as rewarding), (4) Repetition, and (5) Attention. Based on neuroscientific insights, it can be argued that singing may share a larger overlap and requires higher precision in the use of the vocal apparatus than other forms of musical engagement, which – based on this framework – could causally affect non-musical skills like vocal emotion perception. Against this backdrop, the proposition of a null hypothesis is not trivial at all. In fact, it may seem counterintuitive. However, we pursued this approach, because for vocal emotion perception (and other non-musical skills), the OPERA hypothesis has not stood up to rigorous empirical examination (Schellenberg & Lima 2024). Nevertheless, suggestions of causal effects of musical training on non-musical skills in the absence of convincing empirical evidence are still prominent in the literature, which is why studies like ours are important.

We now addressed this in two paragraphs in the introduction:

“*Based on the observation that emotions expressed by voices and by music share a similar acoustic code (Juslin & Laukka, 2003), early theoretical frameworks like the OPERA hypothesis (Patel, 2011) proposed a causal effect of musical training on voice perception skills, if specific conditions are met: an overlap in the neural circuits, precision in auditory-motor demands, as well as involvement of emotion, repetition and attention in the musical activity. However, for vocal emotion recognition, the OPERA hypothesis has not stood up to rigorous empirical examination (Schellenberg & Lima, 2024; Swaminathan & Schellenberg, 2020). Instead, evidence collectively points to the role of auditory sensitivity, which does not seem to be causally linked to formal musical training.*” (page 4)

“*Crucially, singers use their voice for musical expression. This is reflected in vocal performance differences, as for example, singers outperform instrumentalists in voice imitation tasks (Christiner & Reiterer, 2015; Waters et al., 2021). Further, neuroscientific research revealed substantial overlap between the neural circuits involved in the expression and perception of vocal information (Frühholz & Schweinberger, 2021). But how does this relate to the sensitivity in the perception of vocal cues? The abovementioned OPERA hypothesis (Patel, 2011) would predict that singers' high degree of auditory-motor precision and neural overlap would lead to benefits in perception. However, this is not consistently supported by empirical findings (Nikjeh et al., 2009). In fact, […]*” (page 6)

5. **Reviewer #2: In general, the writing is very clear. My one main suggestion is to emphasize the big picture more. If individual differences in vocal emotion recognition are caused by an environmental factor (i.e., music training/performing in this instance), one would expect such differences to depend on the quality and quantity of the environmental intervention, yet none was found.**

Response: Yes, one of the key main points of our story is that there is no evidence for a causal effect of music training on vocal emotion recognition, and this is why we assumed that there should not be an effect of the environmental intervention (in quality and quantity), which is what we found. We adjusted two parts in the introduction, to make this clearer. Importantly, we now brought up an important theoretical framework (the OPERA hypothesis), and outline, why this has not stood up to empirical investigations in the context of vocal emotion recognition. For more details, please refer to our response to Point 4.

6. **I also thought that the ms could be streamlined a bit more by not repeating details that were included in the authors' earlier publication (Nussbaum et al., 2024), but rather referring the reader to the previous article. The authors do this to some extent already. I'm wondering if they could do it more without making reading the earlier paper a prerequisite for understanding the present submission.**

Response: We confess that we were slightly ambivalent about this, partly because there seem to be different views on this between reviews, and partly because we believe it is very important that reading the earlier paper is not the prerequisite for understanding the current one and we feel that a certain level of detail is crucial for that, especially regarding the acoustic manipulations. Nevertheless, we went through the methods section again and checked whether we can reduce redundancy by referring to the previous publication. If the editor or the reviewers have specific ideas for aspects we could leave out without compromising comprehensibility, we are very open to this. So far, we did omit a few less relevant details:

“~~After substantial preprocessing (e.g. manual mapping of time- and frequency anchors in each stimulus),~~” (page 11)

“~~As browser, we recommended Google Chrome, and excluded Safari for technical reasons.~~” (page 12)

7. **Reviewer #3: Major issues: Q1. Transparency and preregistration**
   The study claims to follow open science principles through preregistration. However, the preregistration was submitted on May 3, 2025, more than two years after the original study (Nussbaum, 2024), when some data had already been collected and analyzed. While the current study includes new data collection, the timing of the preregistration, shortly before manuscript submission, raises concerns about whether it qualifies as a true preregistration. Most importantly, this makes it unclear whether the hypotheses were formulated a priori or post hoc. Although I don’t consider this as an attempt to circumvent p-hacking, I believe these issues should be addressed explicitly if the benefits of preregistration and open science are to be fully realized.
   This is my major concern and I would like the authors to elaborate on this, in the manuscript primarily, but of course also in their response.

Response: There seems to be a confusion. The preregistration was submitted/published on June 9, 2023 (link: https://doi.org/10.17605/OSF.IO/76PV5). This was *before* the very first datapoint of the new dataset was collected (data collection started on June 12, 2023). Of course, it was *after* the data of the previous study (Nussbaum 2024) were collected, analyzed, and (almost) published, but this was made fully transparent and explicit in the preregistration throughout (e.g. the description and the explanation of preexisting data). We checked, but we could not find the date May 3 2025 anywhere? There was something like an sever update/migration of OSF earlier this year, maybe thats when some metadata was changed? However, our preregistration was published in June, 2023 and we can also provide some (automatic) e-mail correspondence/notification with a time stamp on request.

We added the specific dates in the manuscript as follows:

“*We specified how we determined our sample size, all data exclusions, all manipulations, and all measures in the associated preregistration (https://doi.org/10.17605/OSF.IO/76PV5), published on June 9, 2023, before we started data collection on June 12, 2023.*” (page 15/16)

8. **Q2. Coding of PROMS Responses**
   The manuscript states:
   "In alignment with the approach by Nussbaum et al. (2024), we recoded responses in the PROMS from 0 to 1 in 0.25 steps starting with the “definitely” correct option down two the “definitely” incorrect option (thus, “don’t know” was always coded with 0.5) and subtracted 0.5 from the final measure"

Should "down two" be "down to"? Is "definitely correct" coded as 0 or as 1? Why is 0.5 added and then subtracted for the “don't know option”. Please explain more clearly.

Response: Good point – “definitely correct” was coded as 1 and “definitely incorrect” was coded as 0. “Down two” should be “down to”. We adjusted this in the manuscript. Thus, each response gets a number between 1 and 0. We then subtracted 0.5. This was mainly done to facilitate interpretability, because this way a positive score indicates more positive/correct ratings, a negative score indicates more incorrect ratings and a score of zero indicates responses at chance level. But of course, the subtraction of a constant does not affect the statistical analyses of group differences. We rephrased this paragraph to make it clearer:

“In alignment with the approach by Nussbaum et al. (2024), we first recoded responses in the PROMS from 1 to 0 in 0.25 steps starting with the “definitely” correct option down to the “definitely” incorrect option (thus, “don’t know” was always coded with 0.5). For the final measure, we then subtracted 0.5, so that a positive score indicates that participants were more correct/confident, a negative score indicates more incorrect/uncertain ratings, and a score of zero indicates responses at chance level. For statistical analyses, we used the averaged performance across trials for each subtest.” (page 15)

9. **Q3. Justification of Priors**
   In Bayesian analysis, the choice of priors is critical. While “default setting of priors” are convenient and often appropriate, they may not reflect the best assumptions for a specific research question. The manuscript does not justify the choice of priors, which is particularly relevant given that Bayes factor outcomes are sensitive to prior assumptions. Without such justification, it is unclear whether the default priors aligns with the expected effect size or the theoretical framework of the study (which anticipates a medium effect, as mentioned in the power analysis). Please provide a rationale for the chosen priors.

Response: For our Bayesian analysis, we oriented ourselves at Neves et al. 2025 (https://doi.org/10.1016/j.cognition.2025.106102) who tested a null hypotheses for a related research questions and with similar sample sizes. For their analysis, they used default priors and it is also argued in Ly et al 2016 (http://dx.doi.org/10.1016/j.jmp.2015.06.004) that default priors are appropriate for the testing of null hypotheses (as the anticipated effect is, in fact, a null effect and not a medium-sized one). We have added an explanatory part to the methods sections:

“*These analyses were conducted in JASP Version 0.19.3 (JASP Team, 2025) using default priors, which have been considered appropriate for testing null hypotheses based on similar sample sizes (Ly et al., 2016; Neves et al., 2025). Further, we ensured that our*

Bayesian inference did not depend critically on the choice of priors by running robustness checks (see data analysis files on OSF)." (page 15)

10. Q4. Sensitivity Analysis
The manuscript does not report whether results are robust to different prior choices. This is a common but important omission in Bayesian reporting. I recommend a sensitivity analysis to demonstrate the reliability of the findings across varying prior assumptions.

Response: In fact, we did include a robustness check in our Bayesian analysis, which we uploaded on OSF but did not mention that in the manuscript. We fixed that now (cf. our response to Point 9).

11. Q5. Rationale for Emotional Categories
The rationale for selecting the emotional categories could be expanded. While the inclusion of two positive (Pleasure, Happiness) and two negative (Fear, Sadness) emotions provides a balanced valence structure, it is unclear whether the authors considered the temporal and psychological distinctions. For example, Pleasure is often immediate and biologically driven, whereas Happiness is more enduring and psychologically constructed. Similarly, Fear is typically short-lived and reactive, while Sadness may be more prolonged. Where these distinctions considered in the design? Could comparisons between short-term and long-term, or comparisons between positive and negative grounded emotions offer additional insights? And if so, as an exploratory analysis.

Response: Concerning our rationale for selecting the emotional categories, please refer to our response to point 3. Regarding the temporal domain, our voice morphing approach ensures that the timing is kept constant across all emotions by holding the time anchors constant. However, that does not imply that the emotional information does unfold similarly over time. It is possible, for example, that the emotional information in happiness is available earlier than in e.g. in sadness. We discussed this possibility in a previous EEG study, because we observed somewhat emotion-specific timing in the electrophysiological responses. This was not the focus of the present study, but we mentioned this work in the discussion for the interested reader (page 29). In short, our inclusion of two positive and two negative emotions of different intensities was guided by the aim to provide balanced stimuli both with respect to valence and intensity. We now clarify this rationale in the methods section, and provide further evidence by reporting additional rating data of emotional intensity of the stimuli.

12. Q6. Inclusion of Big5 Personality Traits
In table 4, the BIG5 personality traits are reported, although they are not linked to

any of the preregistered hypotheses. Was analysis conducted for exploratory purposes? If so, please clarify and consider presenting these results in a separate exploratory section to avoid confusion.

Response: Indeed, the purpose of Table 1 and Table 4 is to provide insights into the comparability of the studied subgroups. This is a cross-sectional design, so there is always a risk of undetected confounds, but at least we wanted to make sure that the groups would not differ with regard to "common candidates" (i.e. socioeconomic status, personality, or affective state), or if so, be transparent about it. In line with our reporting style in our previous study (Nussbaum et al., 2024), we put this at the beginning of the results section to give the readers a good overview over the sample characteristics before we move on to the hypotheses. But we are open to relocate or reframe this sample overview, if this is strongly preferred by the editor or the reviewer. In any case, we added a line:

"First, we checked for important demographic and psychological variables that the two groups were comparable." (page 16)

"Again, we first confirmed that the groups were comparable on important individual variables." (page 24)

13. Q7. Terminology in Hypotheses H5-H7

Some terminology used in hypotheses H5-H7 is not introduced in the methods section. See "General-ME", "Perception Subscale", and "selfrated singing abilities". For many readers, these terms only become clear when reviewing the results, particularly Table 4. Please define these terms earlier in the manuscript.

Response: We tried to improve readability by re-numbering related hypotheses and by slightly rephrasing these hypotheses directly:

"H3c: Averaged-VER and Full-VER are correlated with the general musical sophistication index (General-ME), measured by the Gold-MSI.

H3d: Averaged-VER and Full-VER are correlated with the perception subscale of the Gold-MSI

H3e: Averaged-VER and Full-VER are correlated with self-rated singing abilities Subscale of the Gold-MSI." (page 21)

14. Q8. Hypothesis H9

I did not understand the rationale for hyptothesis H9. Please explain.

Response: We tried to outline the rationale for (the former) Hypothesis 9 (now Hypothesis 5) at the beginning of section 4 (Part III). In principle, we assumed that amateur and professional musicians would perform equal in our emotion recognition task. Thus, we predict a similar null finding as for the contrast between singers and instrumentalists. The rationale behind this was that the link between musicality and vocal emotion perception is not based on the quality or quantity of formal musical training. However, back then, we also considered a number of studies that also found that amateurs were actually better than professionals in a number of non-musical abilities, and we therefore phrased this hypothesis more carefully and included this possibility as well.

Please note that our hypothesis targets F0 and the Full morphs condition only, because we assumed that performance would be equal for the Timbre condition anyway. However, we admit that we missed making this explicit in our preregistration.

We slightly adjusted the paragraph before the hypothesis, to make the rationale more transparent:

"*Mostly, we predicted that amateurs and professional musicians would be comparable regarding vocal emotion recognition. However, as previous evidence reviewed above showed that amateurs can differ from professionals in cognitive abilities which could be linked to emotional sensitivity, we also considered the option that amateurs could be more proficient at making emotional inferences than professionals, reflected in H5 below*."
(page 22/23)

15. Q9. Definition of Singers (and Instrumentalists)
On page 9, singers are defined as individuals who "must not currently be instrumentalists in an orchestra or a band". Did you control for participants' musical history, such as previous experience playing an instrument? If so, how?

Response: This is an important point and one that we picked up for critical reflection in our discussion. While it may seem trivial at first glance, making a clear-cut distinction between singers and instrumentalists is actually very complicated on a practical level to recruit mutually exclusive groups. Thus, a few of our singers did report playing an instrument in the past (details are all on OSF) and many instrumentalists also sing (although not on a regular basis or in an ensemble). One possibility would have been to exclude these individuals but at the potential cost of losing a number of participants, and, consequently, statistical power. Instead, we opted for transparent description of our samples and an honest discussion of this limitation on page 31.

16. Q10. Stimulus Validation
Were the stimuli validated? Figure 5 in Nussbaum (2024) shows confusion rates for different emotions, but perhaps there were specific stimuli that were consistently misclassified? Are misclassifications a result of poor interpretation or poor stimuli representation? For example, in Figure 3, is the poor performance in Timbre condition a result of poor stimuli quality or poor perception? Please elaborate.

Response: Yes, stimuli were validated, for details please refer to our response to point 3 and our additional information now provided in the methods section.

We also calculated confusion rates for the present sample, which can be found in the supplemental materials on OSF (and referenced in the manuscript). Please note that confusion matrices look very similar between all groups (singers vs. instrumentalists; and amateurs vs. professionals vs. non-musicians).

We have now checked misclassifications for the individual stimuli. Out of our 288 stimuli, accuracy was numerically below chance level (ACC < 0.25) for 38 (~13%) stimuli, but most of them were Timbre and F0 morphs, were we deliberately cut down emotion information to investigate the impact of this on emotion recognition performance. So, in our view, it is more informative to check how many Full morphs fall below chance, and that's only 2/96 (~ 2%). Importantly, this pattern is highly similar between singers and instrumentalists. Thus, while the classification of a few select stimuli may have been particularly poor, we would argue that this does not invalidate the conclusions of the study, because this should have affected all groups similarly.

Please note that we did not consider individual stimuli in our statistical analysis as this was not part of our preregistered analysis plan and because this would have compromised our statistical power. However, the data is openly available now, leaving the option for future research.

We can only speculate whether misclassifications are the result of poor interpretation or poor representation, because with the present paradigm, we study both in conjunction. Given that vocal emotion classification is highly comparable across groups, we would rather interpret this pattern as a result of the stimulus material and conclude that timbre is less emotionally informative relative to the pitch contour or the combined information. Please note that we can only make inference about the *relative* importance of cues (i.e. pitch vs timbre). Further, we also refrained from interpreting absolute differences between emotions (i.e. whether happiness is overall better recognized than pleasure or fear), because this depends highly on the stimulus material and response biases of participants, which is why we did not consider this informative.

We mention the possibility of stimulus-specific effects briefly in the discussion now:

"*First, our dataset was limited to four emotions expressed through short pseudowords. Future research should examine the extent to which these findings generalize to other types of vocal material.*" (page 31)

17. Q11. Conclusion and causality

The conclusion that "the ability to recognize emotions in voices is not primarily determined by the type or amount of musical activity, but rather by individual differences in auditory sensitivity" is based on correlations in table 3 and table S10.

First, are both tables including control for musical education? Or is it only controlled for in table S10. Please explain the differences. Second, “determined by” implies a causal relationship. Since this is not an experimental setup where individuals’ auditory sensitivity is manipulated, the causal claim has to be based upon theoretical assumptions (for example by motivating that no other potential causes occur).

Response: Any implication of causality was unintentional and therefore we agree that “determined” was a poor choice of wording, which we now changed to “associated with”. Please refer to our response to point 18, where we list several changes we applied to the manuscript in order to avoid causal language.

In section 3.2. (Data analysis), we report that we controlled all correlations for formal musical education as well (according to our preregistered plan), but this made no difference for the observed patterns. For simplicity, we kept the uncontrolled correlations in the manuscript (S3) and the ones controlled for musical education (S10) in the supplemental materials on OSF.

18. **Reviewer #1: The causal language should be reconsidered, even though it is taken on board from the literature, in part. This is a correlational study.**

Response: Since our study specifically challenges the claim that musical training /activity is causally involved in the link between musicality and vocal emotion perception, any impressions of implied causality by the choice of wording are unintentional. We scanned the manuscript and slightly adjusted the wording in the following parts:

- *“auditory skills promote vocal emotion recognition” ->* “*auditory skills are linked to vocal emotion recognition*” (page 5)
- “*Overall, the few data that are available do not provide clearcut, let alone causal evidence for a specific benefit in vocal emotion recognition by singing over playing an instrument.*” (page 7)
- “*emotion recognition is not related to formal training, but rather to natural differences in auditory sensitivity*” (page 8)
- *“these results suggest that the link between musicality and vocal emotion recognition is driven by individual differences in auditory sensitivity” -> we changed “driven by” to “associated with”* (page 27)
- “*This adds a new perspective to the accumulating evidence that the link between musicality and vocal emotion recognition is predominantly driven by individual differences in natural auditory sensitivity.” -> changed “driven by” to “associated with*” (page 33)

19. **The number of hypotheses seems inappropriate to me, at least. As stated above, commenting on a correlation table is sufficient. The hypotheses 3-7 are not at the same level of the surrounding ones. That should be reflected in the writing.**

Response: This is a valid point. Indeed, in the preregistration, we had first stated the four main hypotheses (which target the comparison of instrumentalists vs singers and amateurs vs. professionals) and then added all the predictions about the correlational relationships. And indeed, these were only included to see if they replicate the previously found patterns from Nussbaum et al. (2024), and are not at the same level as the surrounding ones. To adequately reflect this in the manuscript, we changed hypotheses 3-7 (which are inter-related) into 3a – 3e and adjusted all the numbers accordingly. For transparency, we mentioned in the methods section, that the numbering of hypotheses diverges from the preregistration, but the content does not:

“Please note that the numbering and wording of hypotheses was slightly modified from the preregistration to increase clarity, while not affecting their content.” (page 15)

“*In Part II, we focused on the correlations between auditory sensitivity and vocal emotion recognition. The aim here was to see if the patterns found in Nussbaum et al. (2024) would replicate. Therefore, we formulated the following predictions: 3a [..] 3e […]*” (page 20)

20. **Reviewer #2: The ms is written in APA style but some of the references diverge from APA (e.g., title case for article titles, abbreviated journal names).**

Response: Thank you for pointing that out. We scanned through our references and formatted them in APA-style where necessary.

21. **São Luis Castro should be abbreviated Castro, S. L. in the reference list.**

Response: Thanks for noticing. We fixed this.

22. **p. 30 "...future research should incorporate..." Change "incorporate" to "consider"**

Response: Changed as suggested.

23. **Why/how were the four subtests from the PROMS chosen? These aren't the same as those that are included in the mini-PROMS.**

Response: This is true. We chose these specific subtests because we considered them most informative for our research. Because we are contrasting pitch (F0) and timbre in the voice, we included the subtests that we assumed to capture similar aspects in the musical domain (melody, pitch and timbre) and then added the rhythm task to capture sensitivity to temporal information as well. We included only four subtests to keep overall duration of the study within reasonable limits. We added a small clarification to the manuscript:

“[...] comprised of the four subtests „Melody“, „Pitch“, „Timbre”, and „Rhythm“, which we considered most informative for the present research question.” (page 13)

24. Baldé et al. (2025) now has volume, issue, and page numbers. It came out in the June issue of JEPHPP.

Response: Thank you for this update - we updated the information on this paper accordingly.

25. Reviewer #3: Minor issues: Q12. Clarity in Reporting Simple Main Effects. The reporting of simple main effects following significant interactions is exemplary in format, however, there are some points that require clarification. See for example sentence: "…all pairwise comparisons |ts(77)| = 2.57, ps = .012, ds = 0.28 [0.06 0.49], except for Fear vs. Sadness (|t(87)| = 1.13, p = .261)." This begins with comparisons between F0 and Timbre (conditions), but ends with comparisons Fear vs Sadness (emotions). Please clarify and make sure all relevant comparisons are reported.

Response: In this particular paragraph, we calculated the performance difference between the F0 and the timbre conditions and compared this difference between emotions. We carefully checked the sentence, and we think that this is described here correctly. However, we made this more explicit by adding a little specification and changing e.g. “$M = 0.34$” to “$M_{F0\text{-}Timbre} = 0.34$”. (page 19)

26. Q11 Epsilon HF: in table 2, explain epsilon{HF}, for example in a footnote. (Huynh-Feldt I assume).

Response: We added the following note: “Note. ε_{HF} = Huynh–Feldt (HF) epsilon correction factor in case of violation of the sphericity assumption.” (page 18)

27. Q12. Missing Legend in Table 3: Table 3 lacks a legend. Please add one to clarify the meaning of the columns and values.

Response: Good point. We added a table caption “*Spearman correlations between the PROMS and vocal emotion recognition performance*” and more information in the legend: “*VER_{Avg}: VER performance averaged across all trials, Full-Morphs: VER in the Full Morph condition only, F0-Morphs: VER in the F0 Morph condition only, Timbre-Morphs: VER in the Timbre Morph condition only, $PROMS_{Avg}$: music perception performance averaged across all four subtests of the PROMS (Pitch, Melody, Timbre, and Rhythm).*” (page 22)

28. Q13. Terminology Consistency: The manuscript uses both “Vocal emotion recognition” and “vocal emotion perception”. Please ensure consistent terminology

throughout the manuscript or clarify distinctions if several terms are intentionally used.

Response: We have changed "vocal emotion perception" to "vocal emotion recognition" throughout.

29. **Q14. Comparisons between Professionals and Non-Musicians: In Table 4, some variables (Extraversion, Attention to Detail, Social Skills, Rhythm) show differences between professionals and non-musicians. These comparisons are not reported in the table. I suggest either including them where significant or or explicitly stating that such comparisons are reported elsewhere (Nussbaum, 2024?).**

Response: Indeed, we did not focus on these comparisons, because they are already reported and discussed in the previous publication. This is mentioned in the manuscript under 4.2 Method: "We focused our analysis on the comparison of amateurs with the other two groups, because the comparison of professional musicians and non-musicians is reported in Nussbaum et al. (2024)." To make this more transparent to the readers, we added a note to Table 4 as well: "*For a detailed description and discussion of the differences between professional musicians and non-musicians, please refer to the previous publication (Nussbaum et al., 2024).*" (page 24)

30. **Reviewer #1: The authors do not mention in the Abstract, which emotions were considered or whether a model representing dimensions of emotion recognition was used. Therefore, it is not obvious what they mean by emotion recognition upfront.**

Response: We added this information to the abstract now: *"[…] based on short vocal utterances expressing happiness, pleasure, fear, or sadness." (p2)*

31. **The authors use causal language at times, which is understandable due to the wide-spread use of it in the literature. However, in their own terms, they should rather consider to talk about associations, even when the literature they cite implies causality in contexts such as correlational studies. So, please avoid causal language.**

Response: Please refer to our response to point 18, where we list the detailed changes we made to address this concern.

32. **The findings are relevant as they extend the current knowledge to assert that professional or amateur vocalists or instrumentalists do not differ in their VER performance, neither do personality, or music sophistication play any role. However, this may still not be too surprising in light of previous literature, although this conclusion can now be drawn on a planned study.**
    For example, https://doi.org/10.1037/emo0000770 found that "natural musical

abilities" without training are sufficient to excel in VER, limiting the role of musical training, a study that the authors cite themselves. Note that the sample in that study also was substantial (N=169). This leaves the authors' claim of a substantial sample size as a reason to conduct this study with a bit of a sour taste, albeit the sample size was sufficient for their study. I do not really trust the claim as expressed in the first line of 5.3.

Response: To avoid any potential misunderstanding here, we now specified that we refer to the comparison between singers and instrumentalists in this first sentence of 5.3, when we emphasize the comparison between different subgroups of musicians. This is indeed a distinctive feature of the present study. At the same time, we also explicitly appreciate the substantial sample in Correia et al. (2022):

"To the best of our knowledge, this study is the first substantially powered comparison between different subgroups of musicians (specifically, singers vs. instrumentalists and professionals vs. amateurs)" (page 30)

"Note, however, that links between music perception performance and emotion recognition irrespective of formal musical training have been observed in a substantial Portuguese sample of musically trained and untrained participants who varied widely in their musical skill (Correia et al., 2022)." (page 31)

33. There is ample research by Schellenberg to suggest that musical talent or aptitude rather than musical practice and expertise predicts a range of aspects of auditory functioning, especially in the linguistic or speech domain https://doi.org/10.1037/xlm0000798.

Response: Indeed. Since most of the current insights into music training and non-musical abilities is summarized in their recent and extensive review (Schellenberg & Lima 2024, https://doi.org/10.1146/annurev-psych-032323-051354), we repeatedly refer to this work in our manuscript. We now also include the reference you recommended.

34. The authors report a software error (p.13), but offer no account for how that error could have affected the results.

Response: This is a valid point. We added a small paragraph in the discussion, where we address this issue:

"On a practical note, we must acknowledge the technical randomization error. While in the previous study (Nussbaum et al., 2024) stimuli were drawn only once, as intended, the present code allowed full randomization with duplication and omissions of stimuli. While undoubtedly unfortunate, we are nevertheless confident that this error has not affected our results substantially. First, the classification patterns for different Morph Types and Emotions fully replicate our previous study (cf. Figure 3) and we observed highly similar

correlations between vocal emotion recognition and music perception performance (cf. Table 3). Second, while this issue might have decreased our signal-to-noise ratio, it would not have introduced a specific bias. Thus, we still consider both studies sufficiently comparable." (page 31)

35. **I am struggling with the authors' decision to test 9 hypotheses. I believe that hypothesis-driven approaches are mandatory, but there is a line to be drawn between hypotheses based on previous findings in the literature and ad hoc hypotheses (working hypotheses). Therefore, I would rather like to see the basis for H3-H7 more clearly. Otherwise, a correlation table is sufficient with an associated description of implications. Also, considering an arbitrary N/hypothesis ratio, it turns out that 10 individuals on average account for one hypothesis. Given the massive load of data acquired in the 25 minutes of experimenting (fatigue-effects being ignored), I think we are approaching a grey area in experimental economy. The authors should perhaps run tests whether the quality of responses at the beginning and at the end of the 25-minute period is still the same.**

Response: Please refer to our response to point 19, where we addressed this issue in detail. In short, we had four main hypotheses (after renumbering, now the H1, H2, H4, and H5), which target the comparisons of musical subgroups. The ones about the correlational patterns were more secondary, to check if they would replicate our previous results, and are now numbered more parsimoniously H3a-H3e. We now made this transparent to the readers and also reflected on the importance of hypotheses in the numbering and wording.

We agree that the experiment was not outstandingly short, but we consider 25 minutes well within reasonable limits for this kind of experiment, especially for performing musicians, who are used to stay focused while performing over a longer period of time. In addition, and as reported, we allowed self-paced breaks every 3-5 minutes. Further, one experimenter met with every participant in person for a little feedback talk and although some of them reported that the experiment was exhausting, none of them reported extreme fatigue. We now also compared accuracy between the first and the last 25% of trials, and we observed that participants got better over time (~3.5% increase in accuracy). So, we rather see a practice than a fatigue effect.

36. **Table 5 reports an emotion x morph type interaction. Did I miss a discussion of that effect, or follow-up analyses? Please explain.**

Response: Indeed, we did not follow up on the interaction in Part III. In this part, we pooled the data from Nussbaum et al (2024) und the data from Part I for the analysis. In both datasets separately, we found the strong Emotion x Morph Type interaction (as mentioned in section 2.4.2). Thus, logically, we would find the same interaction in the pooled dataset, but we did not elaborate on this simply because it does not provide new information. To make this transparent to the reader and avoid confusion, we now added a remark:

“*Please note that these effects were already present in the two datasets that entered the current data (reported in Nussbaum et al., 2024 and Part I above). Because they afford no new information, they are not further detailed here.*” (page 26)

Graphical abstract

Click here to access/download
Graphical abstract
Graphical_abstract_musical_subgroups.png

Vocal Emotion Recognition: A Comparison of Singers and Instrumentalists, Amateurs and Professionals

Christine Nussbaum[1,2,5], Jessica Dethloff[1,2], Annett Schirmer[3,2], and Stefan R. Schweinberger[1,2,4]

[1]Department for General Psychology and Cognitive Neuroscience, Friedrich Schiller University Jena, Germany

[2]Voice Research Unit, Friedrich Schiller University, Jena, Germany

[3]Institute of Psychology, University of Innsbruck, Austria

[4]Swiss Center for Affective Sciences, University of Geneva, Switzerland

[5]Lead contact

Correspondence should be addressed to Christine Nussbaum, Department for General Psychology and Cognitive Neuroscience, Friedrich Schiller University Jena, Am Steiger 3/ Haus 1, 07743 Jena, Germany. Tel: +49 (0) 3641 945934, E-Mail: christine.nussbaum@uni-jena.de; or to Stefan R. Schweinberger, Department for General Psychology and Cognitive Neuroscience, Friedrich Schiller University Jena, Am Steiger 3/Haus 1, 07743 Jena, Germany. Tel: +49 (0) 3641 945181, Fax: +49 (0)3641 9 45182, E-Mail: stefan.schweinberger@uni-jena.de. Preprocessed data, analysis scripts and supplemental materials can be found in the associated OSF repository (https://osf.io/ascqx/).

Word count: 8000

Abstract

Musicians outperform non-musicians in vocal emotion recognition, presumably due to differences in auditory sensitivity. However, the current literature is inconclusive regarding differential effects of specific types of musical activity. Given the tight link between expression and perception in vocal emotional communication, we assessed possible effects of whether music is expressed vocally or not, and whether music is performed on an amateur or professional level. Importantly, however, we predicted that vocal emotion recognition would be *unaffected* by the type and amount of musical activity because current evidence argues against a causal role of formal musical education. We compared emotion recognition performance of singers (N= 45) vs. instrumentalists (N = 43) and professional musicians (N = 40) vs. amateurs (N = 88) vs. non-musicians (N = 38), based on short vocal utterances expressing happiness, pleasure, fear, or sadness. Using both frequentist and Bayesian inference, we found the predicted nulleffects for singers vs. instrumentalists, and professionals vs. amateurs. Evidence for an advantage in amateurs vs. non-musicians was inconclusive. Across groups, we replicated the consistent link between vocal emotion recognition and auditory sensitivity. Overall, the current work aligns with the perspective that musicians' advantage in recognizing vocal emotions is rooted in auditory sensitivity, rather than specific types of musical activities or formal training.

Keywords: vocal emotion recognition, singers, instrumentalists, amateurs, musicality

Highlights

- Musicians outperform non-musicians in vocal emotion recognition, but it is currently unclear whether this effect is modulated by the specific type of musical activity
- To address this gap, we compared vocal emotion recognition in singers vs. instrumentalists and in amateurs vs. professional musicians
- In line with our predictions, we found no difference between either of these subgroups, suggesting that vocal emotion recognition is unaffected by the type and amount of musical activity
- Across groups, we replicated the link between natural auditory sensitivity and vocal emotion recognition

1 Introduction: associations between musicality and vocal emotion recognition

The human voice is a prime carrier of emotional information. Therefore, adequate perception of vocal emotions is important for everyday social interaction (Laukka et al., 2016; Schirmer et al., 2025). On average, humans can infer emotions from voices well above chance (Banse & Scherer, 1996; Juslin & Laukka, 2003; Scherer, 2018), but this capacity is subject to great individual variability and seems to be linked to differences in *musicality* (Lima & Castro, 2011). It has been shown repeatedly that musicians outperform non-musician in vocal emotion recognition, although the overall effect size can be considered small to moderate (M. Martins et al., 2021; Nussbaum & Schweinberger, 2021; Schellenberg & Lima, 2024). Several works sought to unravel the potential mechanisms underlying this advantage. Based on the observation that emotions expressed by voices and by music share a similar acoustic code (Juslin & Laukka, 2003), early theoretical frameworks like the OPERA hypothesis (Patel, 2011) proposed a causal effect of musical training on voice perception skills, if specific conditions are met: an **o**verlap in the neural circuits, **p**recision in auditory-motor demands, as well as involvement of **e**motion, **r**epetition and **a**ttention in the musical activity. However, for vocal emotion recognition, the OPERA hypothesis has not stood up to rigorous empirical examination (Schellenberg & Lima, 2024; Swaminathan & Schellenberg, 2020). Instead, evidence collectively points to the role of *auditory sensitivity*, which does not seem to be causally linked to formal musical training. Compared to non-musicians, musicians have on average more fine-grained basic auditory skills, such as pitch and rhythm perception, musical memory, or signal-in-noise discrimination (Baldé et al., 2025; Kraus & Chandrasekaran, 2010), which aids vocal emotion recognition. Importantly, this association between auditory perception and vocal emotion recognition was observed even in the absence of any formal musical training (Correia et al., 2022; Lima et al., 2016; Nussbaum et al., 2024; Thompson et al., 2012; Vigl et al., 2024). Further, Correia et al. (2022) found that the link between musical training and vocal emotion recognition was fully mediated by auditory perception skills. The presumably strongest evidence is provided by a recent randomized-controlled study in school children, which found no causal effects of musical training on vocal emotion recognition performance (Neves et al., 2025). Thus, there is consensus in the literature that the observed performance difference of musicians and non-musicians is due to variations in natural auditory sensitivity rather than the result of formal musical education (Schellenberg & Lima, 2024).

In a previous study, we investigated how musicians' auditory skills are linked to vocal emotion recognition in more detail, by focusing on different auditory cues that transport emotional meaning (Nussbaum et al., 2024). We employed parameter-specific voice morphing to create vocal stimuli that expressed emotion only through fundamental frequency contour (F0), timbre or both. F0 is linked to dynamic pitch variation (also referred to as voice melody), whereas timbre is linked to perceived voice quality (i.e. whether it sounds harsh or gentle). Professional musicians outperformed a group of non-musicians when emotions were expressed by F0 and both cues, but not timbre alone. Thus, musicians seem to be specifically proficient at exploiting melodic patterns to infer vocal emotions.

While the available literature paints a fairly consistent picture regarding the link between musicality and vocal emotion recognition, a key limitation inherent in most studies targeting group differences is that they treat musicians as one uniform group, which is, in fact, highly heterogeneous. On the one hand, there are quantitative differences regarding levels of expertise. On the other hand, there are qualitative differences between musicians linked to a variety of styles, genres, and forms of expression, within the scope of the Western music system and beyond. A particularly interesting distinction in the context of vocal emotions is the one between **singers and instrumentalists**. Singing is arguably the form of musical expression that is most closely related to vocal emotions (Akkermans et al., 2019; Mithen et al., 2006). Another interesting debate revolves around differences between **professional**

musicians and amateurs (Vincenzi et al., 2022). Therefore, the present study targeted these subgroups to explore their capacity for vocal emotion recognition. In what follows, we review current insights and outstanding research gaps, cumulating in the rationale for the present study.

1.1 Singers vs. instrumentalists

Singing and playing an instrument are both fundamental forms of musical expression in humans, but they require very different motor skills and, typically, different amounts of formal musical training (Fisher et al., 2020; Krishnan et al., 2018). Crucially, singers use their voice for musical expression. This is reflected in vocal performance differences, as for example, singers outperform instrumentalists in voice imitation tasks (Christiner & Reiterer, 2015; Waters et al., 2021). Further, neuroscientific research revealed substantial overlap between the neural circuits involved in the expression and perception of vocal information (Frühholz & Schweinberger, 2021). But how does this relate to the sensitivity in the perception of vocal cues? The abovementioned OPERA hypothesis (Patel, 2011) would predict that singers' high degree of auditory-motor precision and neural overlap would lead to benefits in perception. However, this is not consistently supported by empirical findings (Nikjeh et al., 2009). In fact, I. Martins et al. (2022) found no differences in electrophysiological responses to emotional voices between singers and instrumentalists, suggesting similar profiles of auditory processing. Apart from this study, relevant evidence regarding vocal emotion recognition is sparse and inconclusive. Several studies observed correlations between vocal emotion recognition and singing abilities, either self-rated or objectively measured (Correia et al., 2022; Greenspon & Montanaro, 2023; Nussbaum et al., 2024), but all samples comprised both singers and instrumentalists. Intriguingly, a music-intervention study reported that singing may even interfere with vocal emotional processing, while instrument lessons had a positive effect (Thompson et al., 2004). However, the validity of this finding is limited by an extensive drop-out of participants and a small sample size (Schellenberg & Lima, 2024). Overall, the few data that are available do not provide clearcut, let alone causal evidence for a specific benefit in vocal emotion recognition by singing over playing an instrument. We therefore pursued the null hypothesis of there being no such differences. In view of the limitations of previous studies, we recruited a well-powered sample of instrumentalists and singers.

1.2 Amateurs vs. professional musicians

Most musicians start with their formal training in childhood, but when they enter adulthood, they pick different paths: some convert their musical activity into a profession, whereas others pursue another career and keep music as a hobby. Interestingly, these groups seem to display several differences with regard to neurocognitive functioning. While amateurs, unsurprisingly, score lower on musical abilities, they show better cognitive abilities in terms of abstract reasoning than professional musicians (Vincenzi et al., 2022). Amateurs may gain more positive outcomes from their musical activity. Perhaps because it takes up less time in their lives, it is enriching, while at the same time minimizing negative consequences related to, for example, exposure to high amplitude sounds and performance pressure. This also seems to be reflected in general health, which was found to be better in amateurs than professionals (Bonde et al., 2018; Hake et al., 2024; Loveday et al., 2023; Maghiar et al., 2023; Rogenmoser et al., 2018). On a different note, one recent study reported that professionals when compared with amateurs were more likely to experience a state of flow during their musical activity, which is usually considered very enjoyable (Rakei & Bhattacharya, 2024). However, to the best of our knowledge, there have been no comparisons between amateurs and professionals with regard to vocal emotion recognition. This gap is addressed with the present study.

1.3 Rationale of the present study

This study focuses on the comparison between singers and instrumentalists as well as professionals and amateurs, and thus zooms into possible similarities or differences between specific subgroups while using an almost identical protocol as Nussbaum et al. (2024). We report our findings in three parts. For Part I, we recruited an original sample of amateur instrumentalists and singers and compared their vocal emotion recognition, their musical perception performance and self-rated musicality. In Part II, we focused on the correlations between these measures, in order to replicate the link between auditory sensitivity and vocal emotion recognition reported by previous studies. Finally, because all our newly recruited singers and instrumentalists were amateurs, the present study offered the opportunity to compare findings with our previously recruited groups of professional musicians and non-musicians (Nussbaum et al., 2024), which we report in Part III.

As mentioned above, we predicted that singers and instrumentalists as well as professionals and amateurs perform equally well in our vocal emotion recognition task, both for emotions expressed by all available vocal cues, as well as emotions expressed by either F0 or timbre cues. We derived these predictions from previous research suggesting that the neural processing of vocal emotions is comparable in singers and instrumentalists (I. Martins et al., 2022). Moreover, we relied on behavioral evidence that the link between musicality and vocal emotion recognition is not related to formal training, but rather to natural differences in auditory sensitivity (Correia et al., 2022; Neves et al., 2025). This study and its hypotheses have been preregistered (https://doi.org/10.17605/OSF.IO/76PV5).

2 Part I: Comparison of non-professional singers and instrumentalists

2.1 Hypotheses

Regarding the comparison between singers and instrumentalists, we formulated the following hypotheses:

H1: We expect *no* difference between singers and instrumentalists in overall vocal emotion recognition performance.

H2: We expect *no* difference between singers and instrumentalists in vocal emotion recognition performance based on timbre and F0 cues only.

2.2 Method

Note that this is a follow-up to the study reported in Nussbaum et al. (2024). Thus, the stimulus material and the design are almost identical, but we recruited a new sample.

2.2.1 Participants

According to our preregistered plan, we aimed at a sample size of 40 singers (20 male, 20 female) and 40 instrumentalists (20 male, 20 female), because in our previous study, this sample size allowed us to reveal medium-sized group effects (d = 0.56 - 0.81) when we compared professional musicians and non-musicians.

Data were collected in a pseudonymized format from June 2023 to January 2024. All participants were aged between 18 and 54 years and fluent German speakers. Participants provided informed consent before completing the experiment and received compensation in the form of 12.50 € or course credit upon completion. The experiment was in line with the ethical guidelines by the American Psychological Association (APA) and approved by the local ethics committee of the Friedrich Schiller University Jena (Reg.-Nr. FSV 19/045).

In total, we collected data from 94 amateur musicians that were divided into singers and instrumentalists. Recruitment criteria specified that participants had to be non-professional musicians (i.e., they held no music-related academic degree or worked professionally as a musician). Singers were required to be currently active in a choir or another singing group but should not play an instrument actively and regularly (i.e., they must not currently be instrumentalists in an orchestra or a band). Instrumentalists, conversely, were required to be currently active in an orchestra or a band, but they should not engage in singing activities actively and regularly (i.e., they must not currently be in a choir or another singing group).

Singers

We recorded data from 48 singers, of which three were excluded (N = 2 had > 5 % trials of omission, N = 1 had technical issues during stimulus playback). Thus, data from 45 singers were analyzed (22 female, 22 male, 1 diverse, aged 18 to 53 years [M = 27.02, SD = 8.2]). Mean onset age of musical training was 8 years (SD = 3.08, 5 - 20 years). Mean duration of musical training was 10 years (SD = 6.59, 0 – 25 years). Five participants reported that they never had any formal musical training. Two participants reported that they had occasional tinnitus, but without any subjective impairments in daily life.

Instrumentalists

Data from 46 instrumentalists were collected, of which three were excluded. One had technical issues during stimulus playback, one was also active in a choir, one held a master's degree in music science and was therefore regrouped with the professional musicians (see Part III). Thus, data from 43 instrumentalists entered analysis (24 female, 18 male, 1 diverse, aged 18 to 54 years [M = 28.51, SD = 10.64]). Mean onset age of musical training was 7 years (SD = 2.27, 4 - 14 years). Mean duration of musical training was 14 years (SD = 10.00, 0 – 44 years). Four participants reported that they never had any formal musical training, thus were autodidacts. For more details, see Table S1 on OSF.

2.2.2 Stimulus material

As stimulus material, we used parameter-specific voice morphs that express emotional information either through the fundamental frequency contour only (F0), through timbre only (Tbr) or through a combination of both (Full).

For voice morphing, we selected original audio recordings from a database of vocal actor portrayals, comprised of pseudowords (/molen/, /loman/, /belam/) uttered by eight speakers (four male, four female) with expressions of happiness, pleasure, fear, and sadness. We specifically opted for two positive and two negative emotions of different intensities, to balance both valence and arousal. A prior validation study with 20 raters confirmed that the two positive and two negative emotions had different degrees of emotional intensity (happiness > pleasure, $t(19)=9.57$, $p < 0.001$ and fear > sadness, $t(19)=6.58$, $p < 0.001$).

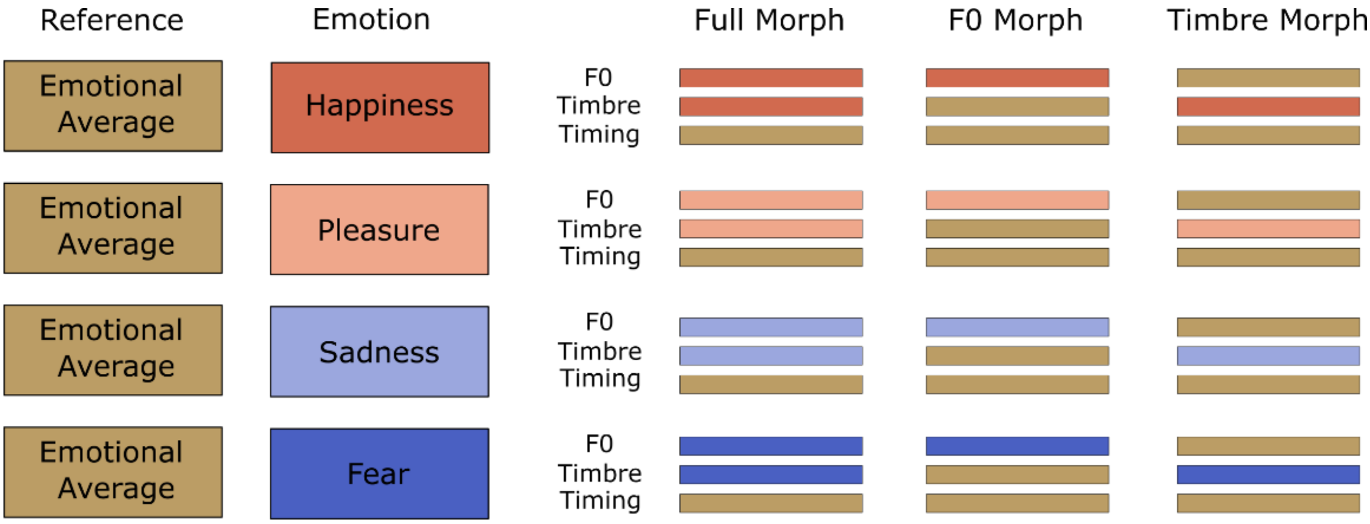
To synthesize the parameter-specific emotional voice morphs, we created morphing trajectories between each emotion and an emotional average of the same speaker and pseudoword, using the Tandem-STRAIGHT software (Kawahara et al., 2013; Kawahara et al., 2008). The averages had been created previously by blending all emotions together and were thus assumed to be uninformative and unbiased with respect to the four emotions of interest. Tandem-STRAIGHT enables voice morphing via weighted interpolation of five independent parameters: (1) F0-contour, (2) timing, (3) spectrum-level, (4) aperiodicity, and (5) spectral frequency; the latter three are summarized as timbre.

We created three types of morphed stimuli (see **Figure 1**). **Full-Morphs** were stimuli with all parameters taken from the emotional version (corresponding to 100% from the emotion and 0% from average), except for the timing parameter, which was taken from the average (corresponding to 0% emotion and 100% average). **F0-Morphs** were stimuli with the F0-contour taken from the emotion, but timbre and timing were taken from the average. **Timbre-Morphs** were stimuli with all timbre parameters taken from the emotion, but F0 and timing from the average. Note that the timing was kept constant in all conditions to allow a pure comparison of F0 vs. timbre. Furthermore, we kept all average stimuli as a further ambiguous reference category. In total, this resulted in 8 (speakers) x 3 (pseudowords) x 4 (emotions) x 3 (morphing conditions) + 24 average (8 speakers x 3 pseudowords) = 312 stimuli (duration M = 780 ms, range 620 to 967 ms, SD = 98 ms). Using PRAAT (Boersma, 2018), we normalized all stimuli to a root-mean-square of 70 dB SPL.

For a more detailed description of the stimulus creation, see Nussbaum et al. (2024) and Kawahara and Skuk (2018). For a summary of acoustic characteristics, see Tables S3 and S4 on OSF.

Figure 1

Morphing matrix for stimuli with averaged voices as reference



Note. Figure reprinted from Nussbaum et al. (2024), Fig 2, page 6

2.2.3 Design

Data were collected online via PsyToolkit (Stoet, 2010, 2017), but after completion of the study all participants met with the experimenter for a short personal debriefing. This was done to increase commitment and conscientiousness for the experiment.

Participants were required to ensure a quiet environment for the duration of the study and use a computer with a physical keyboard and headphones. Prior to the listening tasks, participants could adjust their sound settings to a comfortable sound pressure level.

First, participants entered demographic information, including age, sex, native language, profession, and potential hearing impairments such as tinnitus. They then completed an emotion classification experiment, a test on music perception and several questionnaires on musicality, personality and socioeconomic background. Mean duration of the whole online study was about 75 minutes.

Emotion classification experiment

In the experiment, participants classified vocal emotions as happiness, pleasure, fear, or sadness. Each trial started with a green fixation cross presented for 500 ms. Then the sound was played while a loudspeaker symbol was shown on the screen. Subsequently, a response screen showed the emotion labels and participants could enter their response within a 5000 ms time window starting from voice offset. Responses were entered with the left and right index and middle fingers, with random mapping of response keys to emotion categories for each participant, out of four possible key mappings (see Tables S5 and S6 on OSF). In case of a response omission, the final trial slide (500 ms) prompted participants to respond faster; otherwise, the screen turned black. Then the next trial started.

The 312 stimuli were presented in randomized order in six blocks of 52 trials each, with self-paced breaks in between. Beforehand, participants completed eight practice trials with different stimuli. The experiment was about 25 minutes long. Unfortunately, due to a software error, randomization was sampled with replacements, so that some stimuli were drawn repeatedly and others were omitted. This was in contrast to our previous study, where randomization was sampled without replacement so that each stimulus was drawn exactly once.

Profile of Music Perception Skills (PROMS)

To measure music perception skills, we used a modular version of the Profile of Music Perception Skills (Law & Zentner, 2012; Zentner & Strauss, 2017), comprised of the four subtests „Melody", „Pitch", „Timbre", and „Rhythm", which we considered most informative for the present research question. Participants completed 18 items per subtest, always preceded by one practice trial. Each trial, participants heard a reference stimulus twice followed by a target stimulus. Then, they indicated whether reference and target were the same or different via a 5-point Likert scale with the labels “definitely same”, “maybe same”, “don’t know”, “maybe different”, and “definitely different”. The duration of the PROMS was about 20 minutes.

Questionnaires

After the PROMS, participants completed several questionnaires: the German Version of the Autism Quotient Questionnaire, AQ, (Baron-Cohen et al., 2001; Freitag et al., 2007), a 30-item Personality Inventory measuring the Big-Five domains (Rammstedt et al., 2018), the Goldsmiths Musical Sophistication Index, Gold-MSI, (Müllensiefen et al., 2014) to assess the participants’ degree of self-reported musical skills, additional questions concerning music experience and musical engagement, their socioeconomic background, and the 20-item version of the Positive-Affect-Negative-Affect-Scale, PANAS (Breyer & Bluemke, 2016; Watson et al., 1988).

2.2.4 Data analysis

In line with our preregistered plan, we collapsed data across speakers and pseudowords for analysis. Further, data on emotional averages were excluded because they were not relevant for our hypotheses. Response omissions (~1 %) were treated as errors and participants with more than 5% of such omissions excluded from data analysis. Analyses of Variance (ANOVAs) and correlational analyses were performed using R Version 4.5.0 (R Core Team, 2025). Post-hoc tests were Benjamini-Hochberg corrected where appropriate (Benjamini & Hochberg, 1995).

We complemented these classical frequentist analyses with a Bayesian approach, which – in contrast to null hypothesis significance testing - allows a quantification of evidence for null findings (Rosenfeld & Olson, 2021). These analyses were conducted in JASP Version 0.19.3 (JASP Team, 2025) using default priors, which have been considered appropriate for testing null hypotheses based on similar sample sizes (Ly et al., 2016; Neves et al., 2025). Further, we ensured that our Bayesian inference did not depend critically on the choice of priors by running robustness checks (see data analysis files on OSF). We report the Bayes factor (BF_{10}) as an indicator for the likelihood of the null and alternative hypothesis given the observed data. $BF_{10} > 1$ indicate larger evidence for the alternative hypothesis, $BF_{10} < 1$ indicate larger evidence for the null hypothesis. For example, a $BF_{10} = 3$ means that the alternative hypothesis is three times more likely than the null hypothesis, whereas the reciprocal $BF_{10} = .33$ means that the null hypothesis is three times more likely than the alternative. Following the guidelines by Jarosz and Wiley (2014), we consider values of $BF_{10} = 1\text{-}3$ (.1-.33) as anecdotal, $BF_{10} = 3\text{-}10$ (.33-.10) as moderate, $BF_{10} = 10\text{-}30$ (.10-.03) as strong, $BF_{10} = 30\text{-}100$ (.03-.01) as very strong and $BF_{10} > 100$ (< .01) as decisive evidence for the alternative hypothesis and the reciprocal values in parentheses as respective evidence for the null hypothesis.

In alignment with the approach by Nussbaum et al. (2024), we first recoded responses in the PROMS from 1 to 0 in 0.25 steps starting with the "definitely" correct option down to the "definitely" incorrect option (thus, "don't know" was always coded with 0.5). For the final measure, we then subtracted 0.5, so that a positive score indicates that participants were more correct/confident, a negative score indicates more incorrect/uncertain ratings, and a score of zero indicates responses at chance level. For statistical analyses, we used the averaged performance across trials for each subtest.

2.3 Transparency and openness

We specified how we determined our sample size, all data exclusions, all manipulations, and all measures in the associated preregistration (https://doi.org/10.17605/OSF.IO/76PV5), published on June 9, 2023, thus before we started data collection on June 12, 2023. Please note that the numbering and wording of hypotheses was slightly modified from the preregistration to increase clarity, while not affecting their content. Preprocessed data, analysis scripts and supplemental materials can be found in the associated OSF repository (https://osf.io/ascqx/).

2.4 Results

2.4.1 Demography, musicality, and personality of participants

First, we checked for important demographic and psychological variables that the two groups were comparable. Singers and instrumentalists did not differ significantly in the socioeconomic status assessed via educational level, X^2 (2, N = 88) = 1.06, p = .588; highest academic degree, X^2 (7, N = 88) = 9.06, p = .249, and household income, X^2 (4, N = 88) = 5.23, p = .264 (for more details see Table S2 on OSF). Further, the groups did not differ in age or positive and negative affect (assessed with the PANAS) and were comparable regarding Big Five personality traits and autistic traits. In the Gold-MSI, singers and instrumentalists scored comparatively on the general musicality score, but there were differences on two subfactors: instrumentalists scored higher on the subfactor Formal Education, while singers scored higher on Singing. In the PROMS, both groups performed comparably in all four subtests. Participant characteristics assessed via self-report and music performance in the PROMS are summarized in **Table 1**.

Table 1

Characteristics of participants - Demography, personality, and musicality

	Singers	**Instrumentalists**					
	M (SD)	M (SD)	t	df[a]	p	Cohen's d	
Age	27.02 (8.2)	28.51 (10.6)	-0.73	78.93	.465	-0.17 [-0.61, 0.28]	
PANAS							
positive Affect	3.00 (0.68)	3.11 (0.57)	-0.77	84.78	.446	-0.17 [-0.59, 0.26]	
negative Affect	1.53 (0.47)	1.40 (0.35)	1.49	80.61	.141	0.33 [-0.11, 0.77]	
Big Five							
Openness	4.04 (0.55)	3.99 (0.51)	0.46	85.96	.647	0.10 [-0.32, 0.52]	
Conscientiousness	3.47 (0.69)	3.76 (0.70)	-1.91	85.62	.060	-0.41 [-0.84, 0.02]	
Extraversion	3.21 (0.70)	3.00 (0.73)	1.44	85.3	.155	0.31 [-0.12, 0.74]	
Agreeableness	3.81 (0.57)	4.01 (0.60)	-1.61	85.29	.112	-0.35 [-0.77, 0.08]	
Neuroticism	2.74 (0.77)	2.61 (0.78)	0.80	85.75	.426	0.17 [-0.25, 0.60]	
AQ							
Total	18.2 (6.15)	19.28 (8.55)	-0.68	76.04	.500	-0.16 [-0.60, 0.30]	
Attention to Detail	5.4 (2.33)	5.63 (2.53)	-0.44	84.64	.662	-0.10 [-0.52, 0.33]	
Social	12.8 (5.37)	13.65 (7.53)	-0.61	75.65	.545	-0.14 [-0.59, 0.31]	
Social Skills	2.4 (1.94)	3.09 (2.95)	-1.30	72.01	.200	-0.31 [-0.77, 0.16]	
Communication	2.53 (1.94)	2.44 (2.31)	0.20	82.02	.842	0.04 [-0.39, 0.48]	
Imagination	2.51 (1.75)	2.81 (1.88)	-0.78	84.86	.437	-0.17 [-0.60, 0.26]	
Attention Switching	5.36 (1.91)	5.30 (2.23)	0.12	82.69	.905	0.03 [-0.40, 0.46]	
Gold-MSI							
General ME	4.78 (0.85)	4.75 (0.80)	0.17	85.99	.866	0.04 [-0.39, 0.46]	
Active Engagement	3.83 (0.82)	4.21 (1.13)	-1.79	76.79	.078	-0.41 [-0.86, 0.05]	
Formal Education	**4.39 (1.14)**	**4.95 (0.62)**	-2.85	68.31	.006	-0.69 [-1.18, -0.20]	**
Emotion	5.50 (0.81)	5.60 (0.76)	-0.60	85.99	.549	-0.13 [-0.55, 0.29]	
Singing	**4.98 (0.97)**	**4.19 (1.27)**	3.25	78.56	.002	0.73 [0.27, 1.19]	**
Perception	5.73 (0.82)	5.77 (1.03)	-0.22	80.11	.825	-0.05 [-0.49, 0.39]	
PROMS							
Pitch	0.23 (0.08)	0.24 (0.06)	-0.30	82.92	.766	-0.07 [-0.50, 0.37]	
Melody	0.17 (0.10)	0.14 (0.10)	1.29	85.08	.199	0.28 [-0.15, 0.71]	
Timbre	0.29 (0.08)	0.3 (0.09)	-0.59	86.00	.556	-0.13 [-0.55, 0.30]	
Rhythm	0.31 (0.09)	0.32 (0.09)	-0.56	85.13	.577	-0.12 [-0.55, 0.30]	

Note. Descriptive values show mean ratings for the PANAS (Breyer & Bluemke, 2016), the Big-Five Domains (Rammstedt et al., 2018), and the Gold-MSI (Müllensiefen et al., 2014). AQ score were calculated based on Hoekstra et al. (2008) and Baron-Cohen et al. (2001).

[a] *Note that original degrees of freedom were 86 but were corrected due to unequal variance.*

2.4.2 Emotion classification performance

The mean proportion of correct responses was submitted to an ANOVA with Emotion (Happiness, Pleasure, Fear, and Sadness) and Morph Type (Full, F0, and Timbre) as repeated measures factors and Group (singers and instrumentalists) as a between subject factor (see **Table 2**).

The results revealed main effects of **Emotion** and **Morph Type**, which were qualified by an interaction. Crucially, however, we found no main effects or interactions involving **Group** (see **Figure 2**), which was also confirmed by a Bayesian ANOVA (see Table S7 on OSF). Planned Bayesian analysis revealed moderate evidence for the null effect of group for overall performance (p = .542, BF_{10} = 0.265), as well as for Full (p = .392, BF_{10} = 0.310), F0 (p = .935, BF_{10} = 0.226), and Timbre morphs (p = .555, BF_{10} = 0.262) separately. Thus, we found evidence consistent with our hypotheses H1 and H2.

Table 2

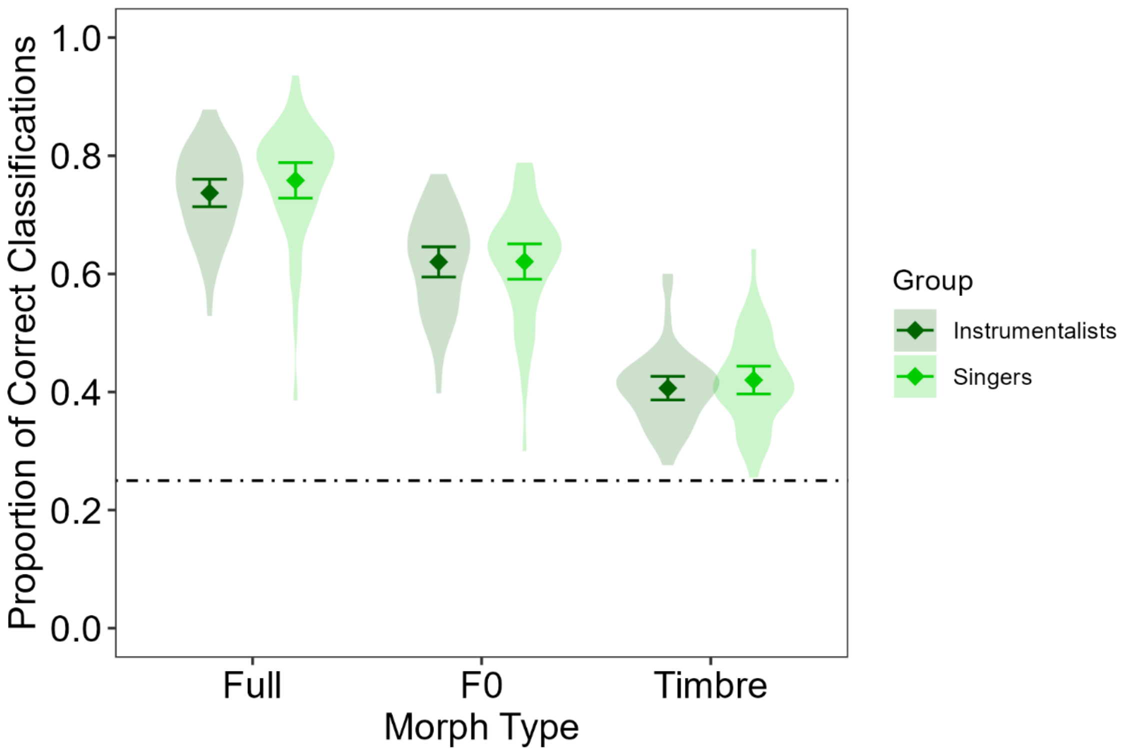
Results of the 4 × 3 × 2 mixed-effects ANOVA on the mean proportion of correct responses

	df1	df2	F	p	Ω_p^2 [95%-CI]	ε_{HF}
Group	1	86	0.38	.542	.00 [.00, .01]	
Emotion	3	258	72.43	<.001	.45 [.36, .52]	
Morph Type	2	172	768.93	<.001	.90 [.87, .92]	.741
Group x Emotion	3	258	2.14	.095	.01 [.00, .04]	
Group x Morph Type	2	172	0.36	.635	.00 [.00, .01]	
Emotion x Morph Type	6	516	22.78	<.001	.20 [.14 .25]	.827
Group x Emotion x Morph Type	6	516	1.33	.249	.00 [.00 .01]	

Note. ε_{HF} *= Huynh–Feldt (HF) epsilon correction factor in case of violation of the sphericity assumption.*

Figure 2

Mean proportion of correct responses per Morph Type separately for singers and instrumentalists

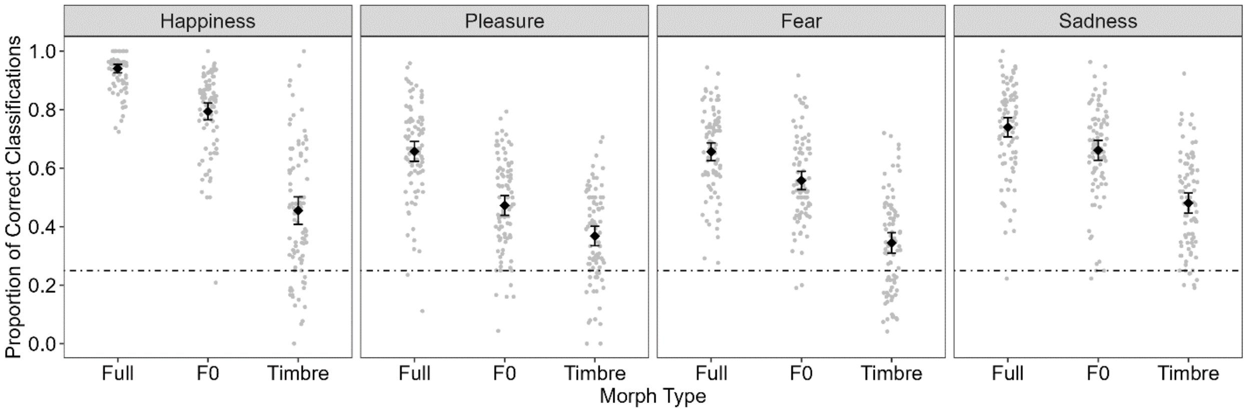


Note. Whiskers represent 95% confidence intervals. Violin plots represent variation of individual participants. The dotted line represents guessing rate at .25.

Follow-up analysis of the Morph Type effect revealed that performance was best in the Full condition (M = 0.75 ± 0.01 SEM), followed by the F0 (M = 0.62 ± 0.01) and then the Timbre condition (M = 0.41 ± 0.01); Full vs. F0: $|t(87)| = 23.28$, $p < .001$, $d = 2.50$ [2.07, 2.92], F0 vs Timbre: $|t(87)| = 21.44$, $p < .001$, $d = 2.30$ [1.90, 2.70], Full vs Timbre: $|t(87)| = 33.94$, $p < .001$, $d = 3.64$ [3.06, 4.21]). This Morph Type main effect was also found for all emotions separately (all Fs(2, 174) > 102.44, $p < .001$), although it differed slightly between emotions, as suggested by the interaction (see **Figure 3,** for all post-hoc tests, refer to the full analysis on OSF). Performance difference between F0 and Timbre was largest for Happiness ($M_{F0\text{-}Timbre}$ = 0.34 ± 0.02 SEM), followed by Fear ($M_{F0\text{-}Timbre}$ = 0.21 ± 0.02), Sadness ($M_{F0\text{-}Timbre}$ = 0.18 ± 0.02), and Pleasure ($M_{F0\text{-}Timbre}$ = 0.10 ± 0.02; all pairwise comparisons $|ts(77)| \geq 2.57$, $ps \leq .012$, $ds \geq 0.28$ [0.06 0.49], except for Fear vs. Sadness ($|t(87)| = 1.13$, p = .261). These effects of Morph Type and Emotion therefore present a full replication of the patterns reported in Nussbaum et al. (2024). For confusion matrices and supplemental analyses, please refer to Figures S1-S3 on OSF.

Figure 3

Mean proportion of correct responses per Emotion and Morph Type



Note. Whiskers represent 95%-confidence intervals. Grey dots represent individual participants' data. The dotted line represents guessing rate at .25.

3 Part II: Correlational analyses

3.1 Hypotheses

In Part II, we focused on the correlations between auditory sensitivity and vocal emotion recognition. The aim here was to see if the patterns found in Nussbaum et al. (2024) would replicate. Therefore, we formulated the following predictions:

Correlations with the PROMS

H3a: Averaged vocal emotion recognition (VER) performance is correlated with averaged music perception performance.

H3b: Full-VER and F0-VER are correlated with melody perception in music

Correlations with the GOLD-MSI:

H3c: Averaged-VER and Full-VER are correlated with the general musical sophistication index (General-ME), measured by the Gold-MSI.

H3d: Averaged-VER and Full-VER are correlated with the perception subscale of the Gold-MSI

H3e: Averaged-VER and Full-VER are correlated with self-rated singing abilities subscale of the Gold-MSI.

3.2 Data analysis

We calculated Spearman correlations between vocal emotion recognition performance and both the PROMS music perception performance and the Gold-MSI self-rated musicality. P-values were adjusted for multiple comparisons using the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995). Following our pre-registered plan, all correlations were controlled for formal musical education as well, but we found that this made no difference to the observed patterns.

3.3 Results

Replicating our previous findings, we obtained a strong correlation between vocal emotion recognition and music perception performance, as measured with the PROMS (Table 3, and Table S10 on OSF). Further, we replicated the specific link between vocal emotion recognition and the melody subtest. Although we did not have a specific hypothesis, we again found a link with Rhythm perception. There were no links between performance in the timbre morph condition and the timbre subtest. Overall, this pattern of correlations almost completely replicates the pattern observed in a previous sample (Nussbaum et al., 2024). Thus, the link between auditory sensitivity and vocal emotion recognition seems to be highly comparable across professional musicians, non-musicians and amateurs.

Table 3

Spearman correlations between the PROMS and vocal emotion recognition performance

	$\text{PROMS}_{\text{Avg}}$	Pitch	Melody	Timbre	Rhythm
VER_{Avg}	**.39 (.002)**	.17 (.142)	**.29 (.017)**	.23 (.050)	**.38 (.002)**
Full-Morphs	**.34 (.005)**	.14 (.219)	**.27 (.022)**	.23 (.050)	**.31 (.009)**
F0-Morphs	**.39 (.002)**	.16 (.186)	**.34 (.005)**	**.24 (.044)**	**.32 (.008)**
Timbre-Morphs	.22 (.054)	.12 (.305)	.08 (.473)	.10 (.352)	**.25 (.039)**

Note. VER = Vocal Emotion Recognition performance. p-values were adjusted for multiple comparisons using the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995). VER_{Avg}: VER performance averaged across all trials, Full-Morphs: VER in the Full Morph condition only, F0-Morphs: VER in the F0 Morph condition only, Timbre-Morphs: VER in the Timbre Morph condition only, $\text{PROMS}_{\text{Avg}}$: music perception performance averaged across all four subtests of the PROMS (Pitch, Melody, Timbre, and Rhythm).

In contrast, we found no links between vocal emotion recognition performance and self-rated musicality, as measured by the Gold-MSI (all correlations ≤ 0.21, *ps* $\geq .051$, both with and without correction for formal musical education, details on Tables S11 and S12 on OSF). Thus, for amateurs, we could not replicate the link with self-rated musical sophistication, perception and singing abilities, which we observed in our previous sample of professional musicians and non-musicians. Therefore, we found evidence for our hypotheses H3a and H3b, but not for hypotheses H3c – H3e.

4 Part III: Comparison of professionals, amateurs and non-musicians

In the third part, we compared amateur musicians to both professional musicians and non-musicians. Mostly, we predicted that amateurs and professional musicians would be comparable regarding vocal emotion recognition. However, as previous evidence reviewed above showed that amateurs can differ from professionals in cognitive abilities which could be linked to emotional sensitivity, we also considered the option that amateurs could be more proficient at making emotional inferences than professionals, reflected in H5 below. Compared to our group of non-musicians (H4), we assumed that amateurs would outperform them when emotions were expressed via full emotion cues and F0 cues only, but not timbre, because this is exactly the pattern we observed for the comparison of professional musicians and non-musicians in Nussbaum et al. (2024).

4.1 Hypotheses

H4: Amateur musicians outperform non-musicians in vocal emotion recognition, in the Full and in the F0 condition.

H5: Amateurs perform equal or better to professional musicians in the Full and the F0 condition.

4.2 Method

For this analysis, we collapsed all participants from Part I into the group of amateur musicians and compared it to the groups of professional musicians and non-musicians reported in Nussbaum et al. (2024). Note that we added one participant to the professional group, because he held a master's degree in music (see Part I), so numbers slightly diverge from the original publication. All professional musicians reported to have a music-related academic degree or a non-academic music qualification. Non-musicians played no instrument or engaged in any other musical activities. For a more detailed description, please refer to Nussbaum et al. (2024).

In total, we analyzed data from 40 professional musicians (20 male, 20 female, aged 20 to 42 years [*M* = 29.6; *SD* = 5.58]), 38 non-musicians (18 male, 20 female, aged 19 to 48 years [*M* = 30.5; *SD* = 6.54]) and 88 amateurs (40 male, 46 female, 2 diverse, aged 18 to 54 years [*M* = 27.8; *SD* = 9.44]).

The stimulus material, design and data analysis were identical to Part I. We focused our analysis on the comparison of amateurs with the other two groups, because the comparison of professional musicians and non-musicians is reported in Nussbaum et al. (2024).

4.3 Results

4.3.1 Demography, musicality, and personality of participants

Again, we first confirmed that the groups were comparable on important individual variables. Professionals, amateurs and non-musicians did not differ in the socioeconomic status assessed via educational level ($\chi^2(6, N = 166) = 11.11$, $p = .085$) and highest academic degree ($\chi^2(16, N = 166) = 24.04$, $p = .089$). However, there were differences regarding household income ($\chi^2(8, N = 166) = 20.19$, $p = .010$, Cramer's V = .25), with amateurs reporting higher household income than professionals and non-musicians.

Participant characteristics are summarized in **Table 4.** For a full report of statistical details, please refer to Tables S13-S21 on OSF. The groups were comparable in age as well as in positive and negative affect (assessed with the PANAS). For the Big Five, analyses of variance revealed group differences for extraversion, with slightly higher levels in professionals than in amateurs. Regarding autistic traits, the three groups did not differ in their overall score, but there were differences on several subscales. In the Gold-MSI, professional musicians scored significantly higher than amateurs on all subscales, which in turn scored higher than non-musicians. This is a pattern (professionals > amateurs > non-musicians) one would expect for self-rated musicality. In the PROMS, professionals outperformed amateurs in the Pitch and Melody subtest, whereas there were no differences in the Timbre and Rhythm subtests. Amateurs performed better than non-musicians in the Pitch, Melody and Rhythm subtest but not in the Timbre subtest. Thus, a clear pattern of professionals > amateurs > non-musicians was only found for melody and pitch.

Table 4

Characteristics of participants - Demography, personality, and musicality

	Professionals		Amateurs		Non-Musicians
	M (SD)		M (SD)		M (SD)
PANAS					
positive Affect	3.32 (0.65)		3.05 (0.63)		3.1 (0.67)
negative Affect	1.69 (0.48)		1.47 (0.42)		1.49 (0.69)
Big Five					
Openness	4.12 (0.50)		4.02 (0.53)		3.81 (0.80)
Conscientiousness	3.49 (0.71)		3.61 (0.70)		3.76 (0.72)
Extraversion	**3.48 (0.66)**	>	**3.11 (0.72)**		3.38 (0.79)
Agreeableness	3.92 (0.57)		3.91 (0.59)		3.75 (0.66)
Neuroticism	2.95 (0.65)		2.69 (0.77)		2.58 (0.82)
AQ					
Total	15.7 (4.98)		18.73 (7.40)		17.58 (6.41)
Attention to Detail	5.43 (2.04)		**5.51 (2.42)**	>	**4.32 (2.01)**
Social	**10.28 (4.70)**	<	**13.22 (6.49)**		13.26 (6.51)
Social Skills	**1.48 (1.68)**	<	**2.74 (2.49)**		2.61 (2.63)
Communication	1.85 (1.61)		2.49 (2.12)		2.39 (1.73)
Imagination	2.18 (1.52)		2.66 (1.81)		2.87 (1.95)
Attention Switching	4.78 (1.91)		5.33 (2.06)		5.39 (1.92)
Gold-MSI					
General ME	**5.68 (0.50)**	>	**4.76 (0.82)**	>	**2.74 (1.07)**
Active Engagement	**4.94 (0.81)**	>	**4.02 (1.00)**	>	**2.95 (1.19)**
Formal Education	**5.95 (0.56)**	**>**	**4.66 (0.96)**	>	**1.71 (0.68)**
Emotion	**5.88 (0.73)**	>	**5.55 (0.78)**	>	**4.95 (1.32)**
Singing	**5.34 (0.83)**	**>**	**4.59 (1.19)**	>	**2.84 (1.26)**
Perception	**6.31 (0.51)**	>	**5.75 (0.92)**	>	**4.22 (1.49)**
PROMS					
Pitch	**0.27 (0.06)**	>	**0.24 (0.07)**	>	**0.18 (0.06)**
Melody	**0.23 (0.08)**	>	**0.16 (0.10)**	>	**0.07 (0.08)**
Timbre	0.32 (0.08)		0.29 (0.08)		0.26 (0.09)
Rhythm	0.33 (0.08)		**0.32 (0.09)**	>	**0.27 (0.08)**

Note. Descriptive values show mean ratings for the PANAS (Breyer & Bluemke, 2016), the Big-Five Domains (Rammstedt et al., 2018), and the Gold-MSI (Müllensiefen et al., 2014). AQ score were calculated based on Hoekstra et al. (2008) and Baron-Cohen et al. (2001).

Comparison signs (“>” or “<”) indicate significant differences. For a full report of statistical details, please refer to OSF. For a detailed description and discussion of the differences between professional musicians and non-musicians, please refer to the previous publication (Nussbaum et al., 2024).

4.3.2 Emotion classification performance

The mean proportion of correct responses was submitted to an ANOVA with Emotion (Happiness, Pleasure, Fear, and Sadness) and Morph Type (Full, F0, and Timbre) as repeated measures factors and Group (professionals, amateurs, and non-musicians) as a between subjects-factor (see **Table 5**). For the parallel Bayesian ANOVA, please refer to S8 on OSF. The results revealed significant main effects of **Emotion** and **Morph Type**, which were qualified by an interaction between **Emotion** and **Morph Type**. Please note that these effects were already present in the two datasets that entered the current data (reported in Nussbaum et al., 2024 and Part I above). Because they afford no new information, they are not further detailed here. Crucially, however, we found no main effect involving **Group** and only a trend for an interaction between **Group and Morph Type**.

Table 5

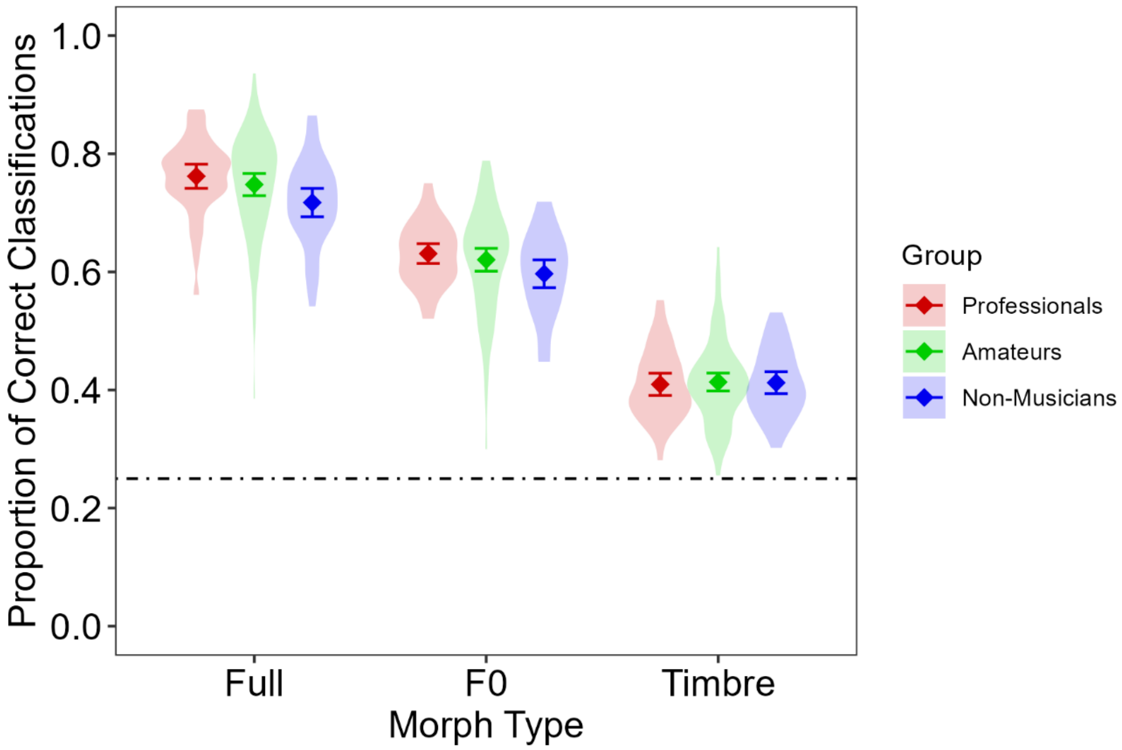
Results of the 4 × 3 × 3 mixed-effects ANOVA on the mean proportion of correct responses

	df1	df2	F	p	Ω_p^2 [95%-CI]	ε_{HF}
Group	2	163	1.96	.144	.01 [.00, .06]	
Emotion	3	489	130.24	<.001	.44 [.38, .49]	
Morph Type	2	326	1357.80	<.001	.89 [.87, .91]	.829
Group x Emotion	6	489	1.17	.322	.00 [.00, .01]	
Group x Morph Type	4	326	2.14	.089	.01 [.00, .04]	.829
Emotion x Morph Type	6	978	40.95	<.001	.20 [.15, .24]	.875
Group x Emotion x Morph Type	12	978	0.74	.688	.00 [.00, .01]	.865

Planned comparisons between **professionals and amateurs** revealed moderate evidence for the null effect of overall performance (p = .473, BF_{10} = 0.238), as well as for Full (p = .322, BF_{10} = 0.286), F0 (p = .435, BF_{10} = .245), and Timbre morphs (p = .840, BF_{10} = 0.205) separately. Planned comparisons between amateurs and non-musicians revealed inconclusive evidence for overall performance (p = .107, BF_{10} = 0.546), as well as for Full (p = .044, BF_{10} = 1.009), and F0 (p = .105, BF_{10} = 0.576) separately. For Timbre morphs, there was moderate evidence for the null effect (p = .975, BF_{10} = 0.205). Thus, we found evidence consistent with H5, but inconclusive evidence regarding H4.

Figure 4

Mean proportion of correct responses per Morph Type separately for professionals, amateurs, and non-musicians



Note. Whiskers represent 95% confidence intervals. Violin plots represent variation of individual participants. The dotted line represents guessing rate at .25.

5 Discussion

In the present study, we shed new light on the link between musicality and vocal emotion recognition, by focusing on different subgroups of musicians. In line with our hypotheses, emotion recognition was found to be comparable between both singers and instrumentalists, and also between amateur musicians and professionals. We further replicated the consistent link between sensitivity towards musical patterns and vocal emotion recognition, although this was reflected in objective performance measures only, but not in self-rated musicality. In total, these results suggest that the link between musicality and vocal emotion recognition is associated with individual differences in auditory sensitivity, which is not tied to a particular type or amount of musical activity. In what follows, we will discuss these findings in more detail.

5.1 Singers vs. instrumentalists

For the present study, we recruited a well-powered sample of singers and instrumentalists, which were comparable with regard to personality, socio-economic background, and objective music perception performance. They only differed in two aspects of self-rated musicality: unsurprisingly, singers scored higher on self-rated singing abilities, and instrumentalists scored higher on self-rated formal education. The latter may be because mastering an instrument to the point where one can play in an ensemble arguably takes more formal training than being able to sing in a choir.

In line with our prediction, the vocal emotion recognition performance of singers and instrumentalists did not differ in any condition, as supported by Bayesian analysis with moderate evidence. This finding does not simply represent a general absence of effects. In fact, we replicated the strong and consistent effect of vocal parameters (manipulated via voice morphing) on emotion recognition, with F0 contour being relatively more informative than timbre, across all emotions, while performance was still best in the Full condition. The fact that this pattern was found to be highly comparable for singers and instrumentalists suggests that they use the same acoustic cues to perceive vocal emotions and do so with comparable efficiency. Note that the effects of the F0 and timbre manipulation on emotion recognition also provided a complete replication of the pattern observed in professional musicians and non-musicians in our previous study (Nussbaum et al., 2024). For an even more detailed reflection on the roles of F0 and timbre for emotion recognition, both on a behavioral and neural level (irrespective of musicality), please also refer to Nussbaum et al. (2022).

Although this is precisely what we predicted, we want to address two observations from the literature which may seem incompatible with our findings: First, our findings diverge from Thompson et al. (2004), who found that singing lessons may interfere with vocal emotional processing. However, this study had several methodological limitations including a very small sample size and large drop-out rates. With the present, substantially powered design, we do not see any evidence for a disruptive effect of active singing. This aligns with the broader literature on protective effects of singing on cognitive and socio-emotional functioning (Moisseinen et al., 2024; Tragantzopoulou & Giannouli, 2025). Second, several studies found correlations between singing abilities and emotion recognition (Correia et al., 2022; Greenspon & Montanaro, 2023; Nussbaum et al., 2024). However, all these studies included both singers and instrumentalists. In fact, it is quite common that instrumentalists have good singing abilities as well, although it is not their preferred form of musical expression. Thus, while we do not argue against a relationship between singing abilities and emotion recognition per se, we lack evidence that emotion recognition is more pronounced in specific groups of musicians, or that this is causally linked to certain musical activities. Instead, this correlation seems to be mediated via natural auditory sensitivity. Note that we failed to find the predicted correlation between self-rated singing abilities and emotion recognition in the present study. This may be attributed to potentially decreased variance in our group of amateurs with regard to self-rated musicality (Gold-MSI). In fact, we did not find any correlations with the Gold-MSI subscales in amateurs, which is in contrast to the pattern we had observed in the sample comprised of professionals and non-musicians (Nussbaum et al., 2024).

5.2 Amateurs compared to professional musicians and non-musicians

As predicted, we found evidence that emotion perception performance does not differ between amateur musicians and professionals, further strengthening the notion that the amount of music training is not a major influence on vocal emotion recognition abilities (Schellenberg & Lima, 2024). We also had hypothesized that amateurs would outperform non-musicians, because professionals outperformed them in our previous study (Nussbaum et al., 2024). This comparison yielded inconclusive evidence. We speculate that the present amateur sample was more heterogenous than the professional one, and as a result, our design may have lacked statistical power to detect potentially small differences between amateurs and non-musicians, despite the substantial sample size. This issue may be resolved in a follow-up study.

In addition to the direct group comparison, we performed a correlational analysis on our sample of amateurs (Part II) which was in parallel to an analysis we had performed previously on professionals and non-musicians (Nussbaum et al., 2024). Importantly, we found a highly similar pattern of correlations between vocal emotion recognition and music perception abilities, especially for melody and rhythm. We therefore conclude that individual differences in music perception abilities play an important role, irrespective of the assignment to any (non)-musical group. This is fully in line with the current literature, which consistently emphasizes the role of auditory sensitivity and argues against a causal effect of training or musical activity on vocal emotion recognition (Neves et al., 2025; Schellenberg & Lima, 2024).

5.3 Directions for future research

To the best of our knowledge, this study is the first substantially powered comparison between different subgroups of musicians (specifically, singers vs. instrumentalists and professionals vs. amateurs) with respect to vocal emotion recognition and therefore closes an open gap in the literature. Nevertheless, the present study has several limitations which deserve consideration. First, our dataset was limited to four emotions expressed through short pseudowords. Future research should examine the extent to which these findings generalize to other types of vocal material. Second, all musicians were socialized in Western music culture and fluent German speakers, therefore findings may not generalize to other music and language backgrounds (Morrison & Demorest, 2009). Note, however, that links between music perception performance and emotion recognition irrespective of formal musical training have been observed in a substantial Portuguese sample of musically trained and untrained participants who varied widely in their musical skill (Correia et al., 2022).

What posed a particular challenge in the present study was the recruitment of mutually exclusive groups of musicians. While the distinction between singers and instrumentalists may seem trivial at first glance, it is actually rare that musicians engage in one of these activities only. Thus, for practical reasons, we recruited participants who self-assigned clearly to one form and made sure that their current activity is either one or the other. Yet, some of the singers reported that they used to play an instrument as well. In a similar vein, the distinction between amateurs and professionals is not fully straightforward and may represent a continuum rather than clear-cut categories. Some individuals pursue music as a profession, but without a formal music degree, others vice versa (Zendel & Alexander, 2020). Some people transition from amateurs to a professional level later in life (Taylor & Hallam, 2011). Thus, while the argument still holds that future research should consider the heterogeneity of musicians, it may be more adequate to pursue variability on a spectrum rather than for distinct groups.

On a practical note, we must acknowledge the technical randomization error. While in the previous study (Nussbaum et al., 2024) stimuli were drawn only once, as intended, the present code allowed full randomization with duplication and omissions of stimuli. While undoubtedly unfortunate, we are nevertheless confident that this error has not affected our results substantially. First, the classification patterns for different Morph Types and Emotions fully replicate our previous study (cf. Figure 3) and we observed highly similar correlations between vocal emotion recognition and music perception performance (cf. Table 3). Second, while this issue might have decreased our signal-to-noise ratio, it would not have introduced a specific bias. Thus, we still consider both studies sufficiently comparable.

While the present behavioral data did not reveal differences between musical subgroups, future research may unravel more fine-grained patterns in the brain. Several studies have reported brain differences in auditory and motor processing between amateurs and professionals (Kleber et al., 2010; Lotze et al., 2003; Oechslin et al., 2013; Papadaki et al., 2023) and singers and instrumentalists (Halwani et al., 2011), but as yet such studies do not provide insights into vocal emotional processing. While we observed differences in electrophysiological responses in professional musicians and non-musicians during vocal emotion recognition (Lehnen et al., 2025; Nussbaum et al., 2023), a meaningful fine-grained comparison between singers and instrumentalists, or amateurs and professionals, would arguably require a much bigger sample.

Finally, with respect to the individual dynamics of the tight neural overlap of expression and perception in vocal communication, we hold that it may be worthwhile to focus on another group of vocal experts: professional voice actors or imitators. This is because such experts, unlike singers, are specifically trained to express emotions in the *spoken* voice. In fact, a recent study found that voice actors show enhanced sensitivity for linguistic voice prosody and enhanced neural tracking for voice pitch compared to non-actors (Kachlicka & Tierney, 2024). Vocal acting usually involves deliberate exaggeration of vocal expressions and may therefore require a fine-grained explicit representation of vocal emotions (Berry & Brown, 2019). While there is research comparing actors and non-actors regarding the expression of emotions (Jürgens et al., 2015), we do not know how this is mirrored in vocal emotion recognition performance and how this group may differ from individuals with singing expertise.

6 Summary and Conclusion

In the present study, we investigated how emotion perception differs in musical subgroups and compared singers and instrumentalists as well as amateurs and professionals. In line with our prediction, we observed no differences between musical subgroups, suggesting that emotion perception abilities are not primarily tied to the type or amount of musical activity. Instead, we replicated the link between vocal emotion recognition and music perception abilities, especially for melodies. This adds a new perspective to the accumulating evidence that the link between musicality and vocal emotion recognition is predominantly associated with individual differences in natural auditory sensitivity.

7 Acknowledgements

This study was conducted by J.D. in partial fulfilment of the requirements for a master's thesis. The original voice recordings that served as a basis for creating our stimulus material were provided by Sascha Frühholz. We thank Hannah Strauß for support with the PROMS. We are grateful to all participants of the study.

8 Conflicts of Interests

The authors declare no competing interests.

9 Declaration of generative AI and AI-assisted technologies in the writing process

No generative AI or AI-assisted technologies were used in the writing process of this manuscript.

10 Credit Author Statement

Christine Nussbaum – Conceptualization, Methodology, Software, Visualization, Formal analysis, Writing - Original Draft, Supervision

Jessica Dethloff - Data collection, Formal analysis, Visualization, Writing - Original Draft

Annett Schirmer – Methodology, Writing - Review & Editing, Supervision

Stefan R. Schweinberger – Conceptualization, Writing - Review & Editing, Supervision

11 Resource availability

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Christine Nussbaum (christine.nussbaum@uni-jena.de).

Data: All raw and preprocessed data have been deposited at the associated OSF repository (https://osf.io/ascqx/) and are publicly available.

Code: All original code for data analysis has been deposited at the associated OSF repository (https://osf.io/ascqx/) and is publicly available.

Additional information: Supplemental figures and tables have been deposited at the associated OSF repository (https://osf.io/ascqx/) and are publicly available.

References

Akkermans, J., Schapiro, R., Müllensiefen, D., Jakubowski, K., Shanahan, D., Baker, D., Busch, V., Lothwesen, K., Elvers, P., Fischinger, T., Schlemmer, K., & Frieler, K. (2019). Decoding emotions in expressive music performances: A multi-lab replication and extension study. *Cognition & Emotion*, *33*(6), 1099–1118. https://doi.org/10.1080/02699931.2018.1541312

Baldé, A. M., Lima, C. F., & Schellenberg, E. G. (2025). Associations between musical expertise and auditory processing. *Journal of Experimental Psychology: Human Perception and Performance*, *51*(6), 747–763. https://doi.org/10.1037/xhp0001312

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614–636. https://doi.org/10.1037/0022-3514.70.3.614

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J.-C., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Berry, M., & Brown, S. (2019). Acting in action: Prosodic analysis of character portrayal during acting. *Journal of Experimental Psychology. General*, *148*(8), 1407–1425. https://doi.org/10.1037/xge0000624

Boersma, P. (2018). Praat: Doing phonetics by computer [Computer program]: Version 6.0.46, retrieved January 2020 from http://www.praat.org/. *Http://www. Praat. Org*.

Bonde, L. O., Juel, K., & Ekholm, O. (2018). Associations between music and health-related outcomes in adult non-musicians, amateur musicians and professional musicians—Results from a nationwide Danish study. *Nordic Journal of Music Therapy*, *27*(4), 262–282. https://doi.org/10.1080/08098131.2018.1439086

Breyer, B., & Bluemke, M. (2016). *Deutsche Version der Positive and Negative Affect Schedule PANAS (GESIS Panel).* https://doi.org/10.6102/zis242

Christiner, M., & Reiterer, S. M. (2015). A Mozart is not a Pavarotti: Singers outperform instrumentalists on foreign accent imitation. *Frontiers in Human Neuroscience*, *9*, 482. https://doi.org/10.3389/fnhum.2015.00482

Correia, A. I., Castro, S. L., MacGregor, C., Müllensiefen, D., Schellenberg, E. G., & Lima, C. F. (2022). Enhanced recognition of vocal emotions in individuals with naturally good musical abilities. *Emotion*, *22*(5), 894–906. https://doi.org/10.1037/emo0000770

Fisher, R. A., Hoult, A. R., & Tucker, W. S. (2020). A Comparison of Facial Muscle Activation for Vocalists and Instrumentalists. *Journal of Music Teacher Education*, *30*(1), 53–64. https://doi.org/10.1177/1057083720947412

Freitag, C. M., Retz-Junginger, P., Retz, W., Seitz, C., Palmason, H., Meyer, J., Rösler, M., & Gontard, A. von (2007). Evaluation der deutschen Version des Autismus-Spektrum-Quotienten (AQ) - die Kurzversion AQ-k. *Zeitschrift Für Klinische Psychologie Und Psychotherapie*, *36*(4), 280–289. https://doi.org/10.1026/1616-3443.36.4.280

Frühholz, S., & Schweinberger, S. R. (2021). Nonverbal auditory communication - Evidence for integrated neural systems for voice signal production and perception. *Progress in Neurobiology*, *199*, 101948. https://doi.org/10.1016/j.pneurobio.2020.101948

Greenspon, E. B., & Montanaro, V. (2023). Singing ability is related to vocal emotion recognition: Evidence for shared sensorimotor processing across speech and music. *Attention, Perception & Psychophysics*, *85*(1), 234–243. https://doi.org/10.3758/s13414-022-02613-0

Hake, R., Kreutz, G., Frischen, U., Schlender, M., Rois-Merz, E., Meis, M., Wagener, K. C., & Siedenburg, K. (2024). A Survey on Hearing Health of Musicians in Professional and Amateur Orchestras. *Trends in Hearing*, *28*, 23312165241293762. https://doi.org/10.1177/23312165241293762

Halwani, G. F., Loui, P., Rüber, T., & Schlaug, G. (2011). Effects of practice and experience on the arcuate fasciculus: Comparing singers, instrumentalists, and non-musicians. *Frontiers in Psychology*, *2*, 156. https://doi.org/10.3389/fpsyg.2011.00156

Hoekstra, R. A., Bartels, M., Cath, D. C., & Boomsma, D. I. (2008). Factor structure, reliability and criterion validity of the Autism-Spectrum Quotient (AQ): A study in Dutch population and patient groups. *Journal of Autism and Developmental Disorders*, *38*(8), 1555–1566. https://doi.org/10.1007/s10803-008-0538-x

Jarosz, A. F., & Wiley, J. (2014). What Are the Odds? A Practical Guide to Computing and Reporting Bayes Factors. *The Journal of Problem Solving*, *7*(1). https://doi.org/10.7771/1932-6246.1167

JASP Team. (2025). *JASP (Version 0.19.3)[Computer software]*. https://jasp-stats.org/

Jürgens, R., Grass, A., Drolet, M., & Fischer, J. (2015). Effect of Acting Experience on Emotion Expression and Recognition in Voice: Non-Actors Provide Better Stimuli than Expected. *Journal of Nonverbal Behavior*, *39*(3), 195–214. https://doi.org/10.1007/s10919-015-0209-5

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814. https://doi.org/10.1037/0033-2909.129.5.770

Kachlicka, M., & Tierney, A. (2024). Voice actors show enhanced neural tracking of pitch, prosody perception, and music perception. *Cortex*, *178*, 213–222. https://doi.org/10.1016/j.cortex.2024.06.016

Kawahara, H., Morise, M., & Skuk, V. G. (2013). Temporally variable multi-aspect N-way morphing based on interference-free speech representations. *IEEE International Conference on Acoustics, Speech and Signal Processing*.

Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., & Banno, H. (2008). TANDEM-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. *IEEE International Conference on Acoustics, Speech and Signal Processing*.

Kawahara, H., & Skuk, V. G. (2018). Voice Morphing. In S. Frühholz, P. Belin, & K. R. Scherer (Eds.), *The Oxford Handbook of Voice Perception* (pp. 684–706). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198743187.013.31

Kleber, B [B.], Veit, R., Birbaumer, N., Gruzelier, J., & Lotze, M. (2010). The brain of opera singers: Experience-dependent changes in functional activation. *Cerebral Cortex*, *20*(5), 1144–1152. https://doi.org/10.1093/cercor/bhp177

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*(8), 599–605. https://doi.org/10.1038/nrn2882

Krishnan, S., Lima, C. F., Evans, S., Chen, S., Guldner, S., Yeff, H., Manly, T., & Scott, S. K. (2018). Beatboxers and Guitarists Engage Sensorimotor Regions Selectively When Listening to the Instruments They can Play. *Cerebral Cortex*, *28*(11), 4063–4079. https://doi.org/10.1093/cercor/bhy208

Laukka, P., Elfenbein, H. A., Thingujam, N. S., Rockstuhl, T., Iraki, F. K., Chui, W., & Althoff, J. (2016). The expression and recognition of emotions in the voice across five nations: A lens model analysis based on acoustic features. *Journal of Personality and Social Psychology*, *111*(5), 686–705. https://doi.org/10.1037/pspi0000066

Law, L. N. C., & Zentner, M. (2012). Assessing musical abilities objectively: Construction and validation of the profile of music perception skills. *PLoS One*, *7*(12), e52508. https://doi.org/10.1371/journal.pone.0052508

Lehnen, J. M., Schweinberger, S. R., & Nussbaum, C. (2025). Vocal Emotion Perception and Musicality-Insights from EEG Decoding. *Sensors (Basel, Switzerland)*, *25*(6). https://doi.org/10.3390/s25061669

Lima, C. F., Brancatisano, O., Fancourt, A., Müllensiefen, D., Scott, S. K., Warren, J. D., & Stewart, L. (2016). Impaired socio-emotional processing in a developmental music disorder. *Scientific Reports*, *6*, 34911. https://doi.org/10.1038/srep34911

Lima, C. F., & Castro, S. L. (2011). Speaking to the trained ear: Musical expertise enhances the recognition of emotions in speech prosody. *Emotion*, *11*(5), 1021–1031. https://doi.org/10.1037/a0024521

Lotze, M., Scheler, G., Tan, H.-R. M., Braun, C., & Birbaumer, N. (2003). The musician's brain: Functional imaging of amateurs and professionals during performance and imagery. *Neuroimage*, *20*(3), 1817–1829. https://doi.org/10.1016/j.neuroimage.2003.07.018

Loveday, C., Musgrave, G., & Gross, S.-A. (2023). Predicting anxiety, depression, and wellbeing in professional and nonprofessional musicians. *Psychology of Music*, *51*(2), 508–522. https://doi.org/10.1177/03057356221096506

Ly, A., Verhagen, J., & Wagenmakers, E.-J. (2016). Harold Jeffreys's default Bayes factor hypothesis tests: Explanation, extension, and application in psychology. *Journal of Mathematical Psychology*, *72*, 19–32. https://doi.org/10.1016/j.jmp.2015.06.004

Maghiar, M. J., Lawrence, B. J., Mulders, W. H., Moyle, T. C., Livings, I., & Jayakody, D. M. P. (2023). Hearing loss and mental health issues in amateur and professional musicians. *Psychology of Music*, *51*(6), 1584–1597. https://doi.org/10.1177/03057356231155970

Martins, I., Lima, C. F., & Pinheiro, A. P. (2022). Enhanced salience of musical sounds in singers and instrumentalists. *Cognitive, Affective, & Behavioral Neuroscience.* Advance online publication. https://doi.org/10.3758/s13415-022-01007-x

Martins, M., Pinheiro, A. P., & Lima, C. F. (2021). Does Music Training Improve Emotion Recognition Abilities? A Critical Review. *Emotion Review*, *13*(3), 199–210. https://doi.org/10.1177/17540739211022035

Mithen, S., Morley, I., Wray, A., Tallerman, M., & Gamble, C. (2006). The Singing Neanderthals: The Origins of Music, Language, Mind and Body by Steven Mithen. London: Weidenfeld & Nicholson, 2005. ISBN 0-297-64317-7 hardback £20 & US$25.2; ix+374 pp. *Cambridge Archaeological Journal*, *16*(1), 97–112. https://doi.org/10.1017/S0959774306000060

Moisseinen, N., Ahveninen, L., Martínez-Molina, N., Sairanen, V., Melkas, S., Kleber, B [Boris], Sihvonen, A. J., & Särkämö, T. (2024). Choir singing is associated with enhanced structural connectivity across the adult lifespan. *Human Brain Mapping*, *45*(7), e26705. https://doi.org/10.1002/hbm.26705

Morrison, S. J., & Demorest, S. M. (2009). Cultural constraints on music perception and cognition. *Progress in Brain Research*, *178*, 67–77. https://doi.org/10.1016/S0079-6123(09)17805-6

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS One*, *9*(2), e89642. https://doi.org/10.1371/journal.pone.0101091

Neves, L., Martins, M., Correia, A. I., Castro, S. L., Schellenberg, E. G., & Lima, C. F. (2025). Does music training improve emotion recognition and cognitive abilities? Longitudinal and correlational evidence from children. *Cognition*, *259*, 106102. https://doi.org/10.1016/j.cognition.2025.106102

Nikjeh, D. A., Lister, J. J., & Frisch, S. A. (2009). The relationship between pitch discrimination and vocal production: Comparison of vocal and instrumental musicians. *The Journal of the Acoustical Society of America*, *125*(1), 328–338. https://doi.org/10.1121/1.3021309

Nussbaum, C., Schirmer, A., & Schweinberger, S. R. (2022). Contributions of fundamental frequency and timbre to vocal emotion perception and their electrophysiological correlates. *Social Cognitive and Affective Neuroscience*, *17*(12), 1145–1154. https://doi.org/10.1093/scan/nsac033

Nussbaum, C., Schirmer, A., & Schweinberger, S. R. (2023). Electrophysiological Correlates of Vocal Emotional Processing in Musicians and Non-Musicians. *Brain Sciences*, *13*(11), 1563. https://doi.org/10.3390/brainsci13111563

Nussbaum, C., Schirmer, A., & Schweinberger, S. R. (2024). Musicality - Tuned to the melody of vocal emotions. *British Journal of Psychology*, *115*(2), 206–225. https://doi.org/10.1111/bjop.12684

Nussbaum, C., & Schweinberger, S. R. (2021). Links Between Musicality and Vocal Emotion Perception. *Emotion Review*, *13*(3), 211–224. https://doi.org/10.1177/17540739211022803

Oechslin, M. S., van de Ville, D., Lazeyras, F., Hauert, C.-A., & James, C. E. (2013). Degree of musical expertise modulates higher order brain functioning. *Cerebral Cortex*, *23*(9), 2213–2224. https://doi.org/10.1093/cercor/bhs206

Papadaki, E., Koustakas, T., Werner, A., Lindenberger, U., Kühn, S., & Wenger, E. (2023). Resting-state functional connectivity in an auditory network differs between aspiring professional and amateur musicians and correlates with performance. *Brain Structure & Function*, *228*(9), 2147–2163. https://doi.org/10.1007/s00429-023-02711-1

Patel, A. D. (2011). Why would Musical Training Benefit the Neural Encoding of Speech? The OPERA Hypothesis. *Front Psychol*, *2*, 142. https://doi.org/10.3389/fpsyg.2011.00142

R Core Team. (2025). *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/

Rakei, A., & Bhattacharya, J. (2024). Professional status matters: Differences in flow proneness between professional and amateur contemporary musicians. *Psychology of Aesthetics, Creativity, and the Arts.* Advance online publication. https://doi.org/10.1037/aca0000674

Rammstedt, B., Danner, D., Soto, C. J., & John, O. P. (2018). Validation of the short and extra-short forms of the Big Five Inventory-2 (BFI-2) and their German adaptations. *European Journal of Psychological Assessment.* Advance online publication. https://doi.org/10.1027/1015-5759/a000481

Rogenmoser, L., Kernbach, J., Schlaug, G., & Gaser, C. (2018). Keeping brains young with making music. *Brain Structure & Function*, *223*(1), 297–305. https://doi.org/10.1007/s00429-017-1491-2

Rosenfeld, J. P., & Olson, J. M. (2021). Bayesian Data Analysis: A Fresh Approach to Power Issues and Null Hypothesis Interpretation. *Applied Psychophysiology and Biofeedback*, *46*(2), 135–140. https://doi.org/10.1007/s10484-020-09502-y

Schellenberg, E. G., & Lima, C. F. (2024). Music Training and Nonmusical Abilities. *Annual Review of Psychology*, *75*, 87–128. https://doi.org/10.1146/annurev-psych-032323-051354

Scherer, K. R. (2018). Acoustic Patterning of Emotion Vocalizations. In S. Frühholz, P. Belin, & K. R. Scherer (Eds.), *The Oxford Handbook of Voice Perception* (pp. 60–92). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198743187.013.4

Schirmer, A., Croy, I., Liebal, K., & Schweinberger, S. R. (2025). Non-verbal effecting - Animal research sheds light on human emotion communication. *Biological Reviews of the Cambridge Philosophical Society*, *100*(1), 245–257. https://doi.org/10.1111/brv.13140

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096–1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*(1), 24–31. https://doi.org/10.1177/0098628316677643

Swaminathan, S., & Schellenberg, E. G. (2020). Musical ability, music training, and language ability in childhood. *J Exp Psychol Learn Mem Cogn*, *46*(12), 2340–2348. https://doi.org/10.1037/xlm0000798

Taylor, A., & Hallam, S. (2011). From leisure to work: amateur musicians taking up instrumental or vocal teaching as a second career. *Music Education Research*, *13*(3), 307–325. https://doi.org/10.1080/14613808.2011.603044

Thompson, W. F., Marin, M. M., & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences*, *109*(46), 19027–19032. https://doi.org/10.1073/pnas.1210344109

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion*, *4*(1), 46–64. https://doi.org/10.1037/1528-3542.4.1.46

Tragantzopoulou, P., & Giannouli, V. (2025). A Song for the Mind: A Literature Review on Singing and Cognitive Health in Aging Populations. *Brain Sciences*, *15*(3). https://doi.org/10.3390/brainsci15030227

Vigl, J., Talamini, F., Strauss, H., & Zentner, M. (2024). Prosodic discrimination skills mediate the association between musical aptitude and vocal emotion recognition ability. *Scientific Reports*, *14*(1), 16462. https://doi.org/10.1038/s41598-024-66889-y

Vincenzi, M., Correia, A. I., Vanzella, P., Pinheiro, A. P., Lima, C. F., & Schellenberg, E. G. (2022). Associations between music training and cognitive abilities: The special case of professional musicians. *Psychology of Aesthetics, Creativity, and the Arts.* Advance online publication. https://doi.org/10.1037/aca0000481

Waters, S., Kanber, E., Lavan, N., Belyk, M., Carey, D., Cartei, V., Lally, C., Miquel, M., & McGettigan, C. (2021). Singers show enhanced performance and neural representation of vocal imitation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *376*(1840), 20200399. https://doi.org/10.1098/rstb.2020.0399

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Zendel, B. R., & Alexander, E. J. (2020). Autodidacticism and Music: Do Self-Taught Musicians Exhibit the Same Auditory Processing Advantages as Formally Trained Musicians? *Frontiers in Neuroscience*, *14*, 752. https://doi.org/10.3389/fnins.2020.00752

Zentner, M., & Strauss, H. (2017). Assessing musical ability quickly and objectively: Development and validation of the Short-PROMS and the Mini-PROMS. *Annals of the New York Academy of Sciences*, *1400*(1), 33–45. https://doi.org/10.1111/nyas.13410

Vocal Emotion Recognition: A Comparison of Singers and Instrumentalists, Amateurs and Professionals

Christine Nussbaum[1,2,5], Jessica Dethloff[1,2], Annett Schirmer[3,2], and Stefan R. Schweinberger[1,2,4]

[1]Department for General Psychology and Cognitive Neuroscience, Friedrich Schiller University Jena, Germany

[2]Voice Research Unit, Friedrich Schiller University, Jena, Germany

[3]Institute of Psychology, University of Innsbruck, Austria

[4]Swiss Center for Affective Sciences, University of Geneva, Switzerland

[5]Lead contact

Correspondence should be addressed to Christine Nussbaum, Department for General Psychology and Cognitive Neuroscience, Friedrich Schiller University Jena, Am Steiger 3/ Haus 1, 07743 Jena, Germany. Tel: +49 (0) 3641 945934, E-Mail: christine.nussbaum@uni-jena.de; or to Stefan R. Schweinberger, Department for General Psychology and Cognitive Neuroscience, Friedrich Schiller University Jena, Am Steiger 3/Haus 1, 07743 Jena, Germany. Tel: +49 (0) 3641 945181, Fax: +49 (0)3641 9 45182, E-Mail: stefan.schweinberger@uni-jena.de. Preprocessed data, analysis scripts and supplemental materials can be found in the associated OSF repository (https://osf.io/ascqx/).

Word count: 8000

Abstract

Musicians outperform non-musicians in vocal emotion recognition, presumably due to differences in auditory sensitivity. However, the current literature is inconclusive regarding differential effects of specific types of musical activity. Given the tight link between expression and perception in vocal emotional communication, we assessed possible effects of whether music is expressed vocally or not, and whether music is performed on an amateur or professional level. Importantly, however, we predicted that vocal emotion recognition would be *unaffected* by the type and amount of musical activity because current evidence argues against a causal role of formal musical education. We compared emotion recognition performance of singers (N= 45) vs. instrumentalists (N = 43) and professional musicians (N = 40) vs. amateurs (N = 88) vs. non-musicians (N = 38), based on short vocal utterances expressing happiness, pleasure, fear, or sadness. Using both frequentist and Bayesian inference, we found the predicted nulleffects for singers vs. instrumentalists, and professionals vs. amateurs. Evidence for an advantage in amateurs vs. non-musicians was inconclusive. Across groups, we replicated the consistent link between vocal emotion recognition and auditory sensitivity. Overall, the current work aligns with the perspective that musicians' advantage in recognizing vocal emotions is rooted in auditory sensitivity, rather than specific types of musical activities or formal training.

Keywords: vocal emotion recognition, singers, instrumentalists, amateurs, musicality

Highlights

- Musicians outperform non-musicians in vocal emotion recognition, but it is currently unclear whether this effect is modulated by the specific type of musical activity
- To address this gap, we compared vocal emotion recognition in singers vs. instrumentalists and in amateurs vs. professional musicians
- In line with our predictions, we found no difference between either of these subgroups, suggesting that vocal emotion recognition is unaffected by the type and amount of musical activity
- Across groups, we replicated the link between natural auditory sensitivity and vocal emotion recognition

1 Introduction: associations between musicality and vocal emotion recognition

The human voice is a prime carrier of emotional information. Therefore, adequate perception of vocal emotions is important for everyday social interaction (Laukka et al., 2016; Schirmer et al., 2025). On average, humans can infer emotions from voices well above chance (Banse & Scherer, 1996; Juslin & Laukka, 2003; Scherer, 2018), but this capacity is subject to great individual variability and seems to be linked to differences in *musicality* (Lima & Castro, 2011). It has been shown repeatedly that musicians outperform non-musician in vocal emotion recognition, although the overall effect size can be considered small to moderate (M. Martins et al., 2021; Nussbaum & Schweinberger, 2021; Schellenberg & Lima, 2024). Several works sought to unravel the potential mechanisms underlying this advantage. Based on the observation that emotions expressed by voices and by music share a similar acoustic code (Juslin & Laukka, 2003), early theoretical frameworks like the OPERA hypothesis (Patel, 2011) proposed a causal effect of musical training on voice perception skills, if specific conditions are met: an **o**verlap in the neural circuits, **p**recision in auditory-motor demands, as well as involvement of **e**motion, **r**epetition and **a**ttention in the musical activity. However, for vocal emotion recognition, the OPERA hypothesis has not stood up to rigorous empirical examination (Schellenberg & Lima, 2024; Swaminathan & Schellenberg, 2020). Instead, evidence collectively points to the role of *auditory sensitivity*, which does not seem to be causally linked to formal musical training. Compared to non-musicians, musicians have on average more fine-grained basic auditory skills, such as pitch and rhythm perception, musical memory, or signal-in-noise discrimination (Baldé et al., 2025; Kraus & Chandrasekaran, 2010), which aids vocal emotion recognition. Importantly, this association between auditory perception and vocal emotion recognition was observed even in the absence of any formal musical training (Correia et al., 2022; Lima et al., 2016; Nussbaum et al., 2024; Thompson et al., 2012; Vigl et al., 2024). Further, Correia et al. (2022) found that the link between musical training and vocal emotion recognition was fully mediated by auditory perception skills. The presumably strongest evidence is provided by a recent randomized-controlled study in school children, which found no causal effects of musical training on vocal emotion recognition performance (Neves et al., 2025). Thus, there is consensus in the literature that the observed performance difference of musicians and non-musicians is due to variations in natural auditory sensitivity rather than the result of formal musical education (Schellenberg & Lima, 2024).

In a previous study, we investigated how musicians' auditory skills are linked to vocal emotion recognition in more detail, by focusing on different auditory cues that transport emotional meaning (Nussbaum et al., 2024). We employed parameter-specific voice morphing to create vocal stimuli that expressed emotion only through fundamental frequency contour (F0), timbre or both. F0 is linked to dynamic pitch variation (also referred to as voice melody), whereas timbre is linked to perceived voice quality (i.e. whether it sounds harsh or gentle). Professional musicians outperformed a group of non-musicians when emotions were expressed by F0 and both cues, but not timbre alone. Thus, musicians seem to be specifically proficient at exploiting melodic patterns to infer vocal emotions.

While the available literature paints a fairly consistent picture regarding the link between musicality and vocal emotion recognition, a key limitation inherent in most studies targeting group differences is that they treat musicians as one uniform group, which is, in fact, highly heterogeneous. On the one hand, there are quantitative differences regarding levels of expertise. On the other hand, there are qualitative differences between musicians linked to a variety of styles, genres, and forms of expression, within the scope of the Western music system and beyond. A particularly interesting distinction in the context of vocal emotions is the one between **singers and instrumentalists**. Singing is arguably the form of musical expression that is most closely related to vocal emotions (Akkermans et al., 2019; Mithen et al., 2006). Another interesting debate revolves around differences between **professional**

musicians and amateurs (Vincenzi et al., 2022). Therefore, the present study targeted these subgroups to explore their capacity for vocal emotion recognition. In what follows, we review current insights and outstanding research gaps, cumulating in the rationale for the present study.

1.1 Singers vs. instrumentalists

Singing and playing an instrument are both fundamental forms of musical expression in humans, but they require very different motor skills and, typically, different amounts of formal musical training (Fisher et al., 2020; Krishnan et al., 2018). Crucially, singers use their voice for musical expression. This is reflected in vocal performance differences, as for example, singers outperform instrumentalists in voice imitation tasks (Christiner & Reiterer, 2015; Waters et al., 2021). Further, neuroscientific research revealed substantial overlap between the neural circuits involved in the expression and perception of vocal information (Frühholz & Schweinberger, 2021). But how does this relate to the sensitivity in the perception of vocal cues? The abovementioned OPERA hypothesis (Patel, 2011) would predict that singers' high degree of auditory-motor precision and neural overlap would lead to benefits in perception. However, this is not consistently supported by empirical findings (Nikjeh et al., 2009). In fact, I. Martins et al. (2022) found no differences in electrophysiological responses to emotional voices between singers and instrumentalists, suggesting similar profiles of auditory processing. Apart from this study, relevant evidence regarding vocal emotion recognition is sparse and inconclusive. Several studies observed correlations between vocal emotion recognition and singing abilities, either self-rated or objectively measured (Correia et al., 2022; Greenspon & Montanaro, 2023; Nussbaum et al., 2024), but all samples comprised both singers and instrumentalists. Intriguingly, a music-intervention study reported that singing may even interfere with vocal emotional processing, while instrument lessons had a positive effect (Thompson et al., 2004). However, the validity of this finding is limited by an extensive drop-out of participants and a small sample size (Schellenberg & Lima, 2024). Overall, the few data that are available do not provide clearcut, let alone causal evidence for a specific benefit in vocal emotion recognition by singing over playing an instrument. We therefore pursued the null hypothesis of there being no such differences. In view of the limitations of previous studies, we recruited a well-powered sample of instrumentalists and singers.

1.2 Amateurs vs. professional musicians

Most musicians start with their formal training in childhood, but when they enter adulthood, they pick different paths: some convert their musical activity into a profession, whereas others pursue another career and keep music as a hobby. Interestingly, these groups seem to display several differences with regard to neurocognitive functioning. While amateurs, unsurprisingly, score lower on musical abilities, they show better cognitive abilities in terms of abstract reasoning than professional musicians (Vincenzi et al., 2022). Amateurs may gain more positive outcomes from their musical activity. Perhaps because it takes up less time in their lives, it is enriching, while at the same time minimizing negative consequences related to, for example, exposure to high amplitude sounds and performance pressure. This also seems to be reflected in general health, which was found to be better in amateurs than professionals (Bonde et al., 2018; Hake et al., 2024; Loveday et al., 2023; Maghiar et al., 2023; Rogenmoser et al., 2018). On a different note, one recent study reported that professionals when compared with amateurs were more likely to experience a state of flow during their musical activity, which is usually considered very enjoyable (Rakei & Bhattacharya, 2024). However, to the best of our knowledge, there have been no comparisons between amateurs and professionals with regard to vocal emotion recognition. This gap is addressed with the present study.

1.3 Rationale of the present study

This study focuses on the comparison between singers and instrumentalists as well as professionals and amateurs, and thus zooms into possible similarities or differences between specific subgroups while using an almost identical protocol as Nussbaum et al. (2024). We report our findings in three parts. For Part I, we recruited an original sample of amateur instrumentalists and singers and compared their vocal emotion recognition, their musical perception performance and self-rated musicality. In Part II, we focused on the correlations between these measures, in order to replicate the link between auditory sensitivity and vocal emotion recognition reported by previous studies. Finally, because all our newly recruited singers and instrumentalists were amateurs, the present study offered the opportunity to compare findings with our previously recruited groups of professional musicians and non-musicians (Nussbaum et al., 2024), which we report in Part III.

As mentioned above, we predicted that singers and instrumentalists as well as professionals and amateurs perform equally well in our vocal emotion recognition task, both for emotions expressed by all available vocal cues, as well as emotions expressed by either F0 or timbre cues. We derived these predictions from previous research suggesting that the neural processing of vocal emotions is comparable in singers and instrumentalists (I. Martins et al., 2022). Moreover, we relied on behavioral evidence that the link between musicality and vocal emotion recognition is not related to formal training, but rather to natural differences in auditory sensitivity (Correia et al., 2022; Neves et al., 2025). This study and its hypotheses have been preregistered (https://doi.org/10.17605/OSF.IO/76PV5).

2 Part I: Comparison of non-professional singers and instrumentalists

2.1 Hypotheses

Regarding the comparison between singers and instrumentalists, we formulated the following hypotheses:

H1: We expect *no* difference between singers and instrumentalists in overall vocal emotion recognition performance.

H2: We expect *no* difference between singers and instrumentalists in vocal emotion recognition performance based on timbre and F0 cues only.

2.2 Method

Note that this is a follow-up to the study reported in Nussbaum et al. (2024). Thus, the stimulus material and the design are almost identical, but we recruited a new sample.

2.2.1 Participants

According to our preregistered plan, we aimed at a sample size of 40 singers (20 male, 20 female) and 40 instrumentalists (20 male, 20 female), because in our previous study, this sample size allowed us to reveal medium-sized group effects ($d = 0.56 - 0.81$) when we compared professional musicians and non-musicians.

Data were collected in a pseudonymized format from June 2023 to January 2024. All participants were aged between 18 and 54 years and fluent German speakers. Participants provided informed consent before completing the experiment and received compensation in the form of 12.50 € or course credit upon completion. The experiment was in line with the ethical guidelines by the American Psychological Association (APA) and approved by the local ethics committee of the Friedrich Schiller University Jena (Reg.-Nr. FSV 19/045).

In total, we collected data from 94 amateur musicians that were divided into singers and instrumentalists. Recruitment criteria specified that participants had to be non-professional musicians (i.e., they held no music-related academic degree or worked professionally as a musician). Singers were required to be currently active in a choir or another singing group but should not play an instrument actively and regularly (i.e., they must not currently be instrumentalists in an orchestra or a band). Instrumentalists, conversely, were required to be currently active in an orchestra or a band, but they should not engage in singing activities actively and regularly (i.e., they must not currently be in a choir or another singing group).

Singers

We recorded data from 48 singers, of which three were excluded (N = 2 had > 5 % trials of omission, N = 1 had technical issues during stimulus playback). Thus, data from 45 singers were analyzed (22 female, 22 male, 1 diverse, aged 18 to 53 years [*M* = 27.02, *SD* = 8.2]). Mean onset age of musical training was 8 years (*SD* = 3.08, 5 - 20 years). Mean duration of musical training was 10 years (*SD* = 6.59, 0 – 25 years). Five participants reported that they never had any formal musical training. Two participants reported that they had occasional tinnitus, but without any subjective impairments in daily life.

Instrumentalists

Data from 46 instrumentalists were collected, of which three were excluded. One had technical issues during stimulus playback, one was also active in a choir, one held a master's degree in music science and was therefore regrouped with the professional musicians (see Part III). Thus, data from 43 instrumentalists entered analysis (24 female, 18 male, 1 diverse, aged 18 to 54 years [*M* = 28.51, *SD* = 10.64]). Mean onset age of musical training was 7 years (*SD* = 2.27, 4 - 14 years). Mean duration of musical training was 14 years (*SD* = 10.00, 0 – 44 years). Four participants reported that they never had any formal musical training, thus were autodidacts. For more details, see Table S1 on OSF.

2.2.2 Stimulus material

As stimulus material, we used parameter-specific voice morphs that express emotional information either through the fundamental frequency contour only (F0), through timbre only (Tbr) or through a combination of both (Full).

For voice morphing, we selected original audio recordings from a database of vocal actor portrayals, comprised of pseudowords (/molen/, /loman/, /belam/) uttered by eight speakers (four male, four female) with expressions of happiness, pleasure, fear, and sadness. We specifically opted for two positive and two negative emotions of different intensities, to balance both valence and arousal. A prior validation study with 20 raters confirmed that the two positive and two negative emotions had different degrees of emotional intensity (happiness > pleasure, $t(19)=9.57$, $p < 0.001$ and fear > sadness, $t(19)=6.58$, $p < 0.001$).

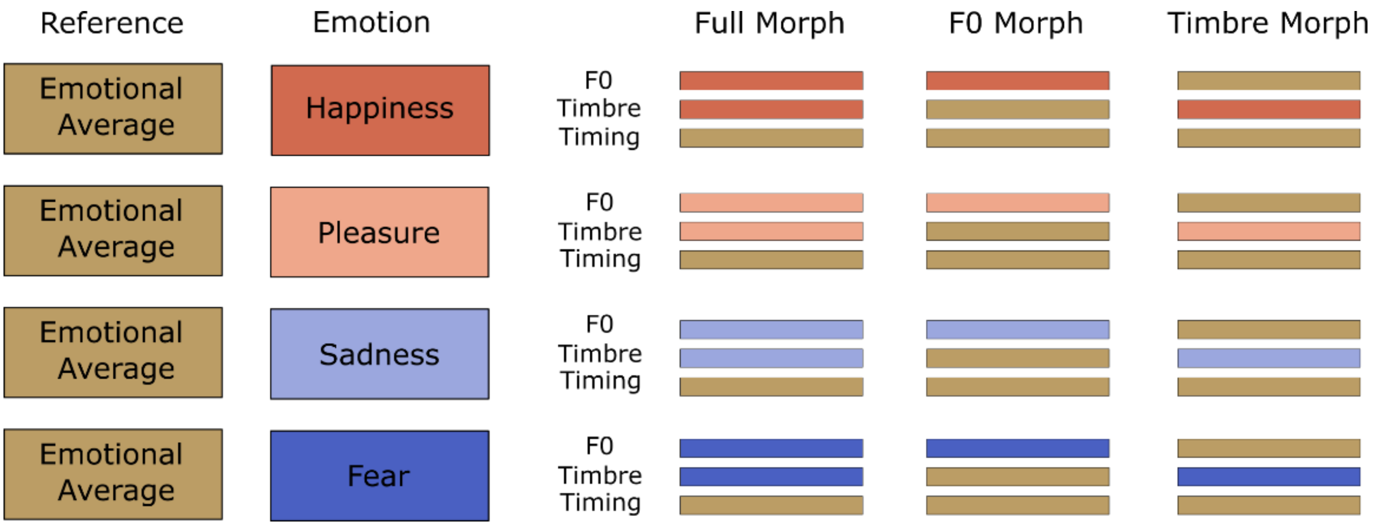
To synthesize the parameter-specific emotional voice morphs, we created morphing trajectories between each emotion and an emotional average of the same speaker and pseudoword, using the Tandem-STRAIGHT software (Kawahara et al., 2013; Kawahara et al., 2008). The averages had been created previously by blending all emotions together and were thus assumed to be uninformative and unbiased with respect to the four emotions of interest. ~~After substantial preprocessing (e.g. manual mapping of time- and frequency anchors in each stimulus),~~ Tandem-STRAIGHT enables voice morphing via weighted interpolation of five independent parameters: (1) F0-contour, (2) timing, (3) spectrum-level, (4) aperiodicity, and (5) spectral frequency; the latter three are summarized as timbre.

We created three types of morphed stimuli (see **Figure 1**). **Full-Morphs** were stimuli with all parameters taken from the emotional version (corresponding to 100% from the emotion and 0% from average), except for the timing parameter, which was taken from the average (corresponding to 0% emotion and 100% average). **F0-Morphs** were stimuli with the F0-contour taken from the emotion, but timbre and timing were taken from the average. **Timbre-Morphs** were stimuli with all timbre parameters taken from the emotion, but F0 and timing from the average. Note that the timing was kept constant in all conditions to allow a pure comparison of F0 vs. timbre. Furthermore, we kept all average stimuli as a further ambiguous reference category. In total, this resulted in 8 (speakers) x 3 (pseudowords) x 4 (emotions) x 3 (morphing conditions) + 24 average (8 speakers x 3 pseudowords) = 312 stimuli (duration M = 780 ms, range 620 to 967 ms, SD = 98 ms). Using PRAAT (Boersma, 2018), we normalized all stimuli to a root-mean-square of 70 dB SPL.

For a more detailed description of the stimulus creation, see Nussbaum et al. (2024) and Kawahara and Skuk (2018). For a summary of acoustic characteristics, see Tables S3 and S4 on OSF.

Figure 1

Morphing matrix for stimuli with averaged voices as reference



Note. Figure reprinted from Nussbaum et al. (2024), Fig 2, page 6

2.2.3 Design

Data were collected online via PsyToolkit (Stoet, 2010, 2017), but after completion of the study all participants met with the experimenter for a short personal debriefing. This was done to increase commitment and conscientiousness for the experiment.

Participants were required to ensure a quiet environment for the duration of the study and use a computer with a physical keyboard and headphones. ~~As browser, we recommended Google Chrome, and excluded Safari for technical reasons.~~ Prior to the listening tasks, participants could adjust their sound settings to a comfortable sound pressure level.

First, participants entered demographic information, including age, sex, native language, profession, and potential hearing impairments such as tinnitus. They then completed an emotion classification experiment, a test on music perception and several questionnaires on musicality, personality and socioeconomic background. Mean duration of the whole online study was about 75 minutes.

Emotion classification experiment

In the experiment, participants classified vocal emotions as happiness, pleasure, fear, or sadness. Each trial started with a green fixation cross presented for 500 ms. Then the sound was played while a loudspeaker symbol was shown on the screen. Subsequently, a response screen showed the emotion labels and participants could enter their response within a 5000 ms time window starting from voice offset. Responses were entered with the left and right index and middle fingers, with random mapping of response keys to emotion categories for each participant, out of four possible key mappings (see Tables S5 and S6 on OSF). In case of a response omission, the final trial slide (500 ms) prompted participants to respond faster; otherwise, the screen turned black. Then the next trial started.

The 312 stimuli were presented in randomized order in six blocks of 52 trials each, with self-paced breaks in between. Beforehand, participants completed eight practice trials with different stimuli. The experiment was about 25 minutes long. Unfortunately, due to a software error, randomization was sampled with replacements, so that some stimuli were drawn repeatedly and others were omitted. This was in contrast to our previous study, where randomization was sampled without replacement so that each stimulus was drawn exactly once.

Profile of Music Perception Skills (PROMS)

To measure music perception skills, we used a modular version of the Profile of Music Perception Skills (Law & Zentner, 2012; Zentner & Strauss, 2017), comprised of the four subtests „Melody“, „Pitch“, „Timbre”, and „Rhythm“, which we considered most informative for the present research question. Participants completed 18 items per subtest, always preceded by one practice trial. Each trial, participants heard a reference stimulus twice followed by a target stimulus. Then, they indicated whether reference and target were the same or different via a 5-point Likert scale with the labels “definitely same”, “maybe same”, “don’t know”, “maybe different”, and “definitely different”. The duration of the PROMS was about 20 minutes.

Questionnaires

After the PROMS, participants completed several questionnaires: the German Version of the Autism Quotient Questionnaire, AQ, (Baron-Cohen et al., 2001; Freitag et al., 2007), a 30-item Personality Inventory measuring the Big-Five domains (Rammstedt et al., 2018), the Goldsmiths Musical Sophistication Index, Gold-MSI, (Müllensiefen et al., 2014) to assess the participants’ degree of self-reported musical skills, additional questions concerning music experience and musical engagement, their socioeconomic background, and the 20-item version of the Positive-Affect-Negative-Affect-Scale, PANAS (Breyer & Bluemke, 2016; Watson et al., 1988).

2.2.4 Data analysis

In line with our preregistered plan, we collapsed data across speakers and pseudowords for analysis. Further, data on emotional averages were excluded because they were not relevant for our hypotheses. Response omissions (~1 %) were treated as errors and participants with more than 5% of such omissions excluded from data analysis. Analyses of Variance (ANOVAs) and correlational analyses were performed using R Version 4.5.0 (R Core Team, 2025). Post-hoc tests were Benjamini-Hochberg corrected where appropriate (Benjamini & Hochberg, 1995).

We complemented these classical frequentist analyses with a Bayesian approach, which – in contrast to null hypothesis significance testing - allows a quantification of evidence for null findings (Rosenfeld & Olson, 2021). These analyses were conducted in

JASP Version 0.19.3 (JASP Team, 2025) using default priors, which have been considered appropriate for testing null hypotheses based on similar sample sizes (Ly et al., 2016; Neves et al., 2025). Further, we ensured that our Bayesian inference did not depend critically on the choice of priors by running robustness checks (see data analysis files on OSF). We report the Bayes factor (BF_{10}) as an indicator for the likelihood of the null and alternative hypothesis given the observed data. $BF_{10} > 1$ indicate larger evidence for the alternative hypothesis, $BF_{10} < 1$ indicate larger evidence for the null hypothesis. For example, a $BF_{10} = 3$ means that the alternative hypothesis is three times more likely than the null hypothesis, whereas the reciprocal $BF_{10} = .33$ means that the null hypothesis is three times more likely than the alternative. Following the guidelines by Jarosz and Wiley (2014), we consider values of BF_{10} = 1-3 (.1-.33) as anecdotal, BF_{10} = 3-10 (.33-.10) as moderate, BF_{10} = 10-30 (.10-.03) as strong, BF_{10} = 30-100 (.03-.01) as very strong and $BF_{10} > 100$ ($< .01$) as decisive evidence for the alternative hypothesis and the reciprocal values in parentheses as respective evidence for the null hypothesis.

In alignment with the approach by Nussbaum et al. (2024), we first recoded responses in the PROMS from 1 to 0 in 0.25 steps starting with the "definitely" correct option down to the "definitely" incorrect option (thus, "don't know" was always coded with 0.5). For the final measure, we then subtracted 0.5, so that a positive score indicates that participants were more correct/confident, a negative score indicates more incorrect/uncertain ratings, and a score of zero indicates responses at chance level. For statistical analyses, we used the averaged performance across trials for each subtest.

2.3 Transparency and openness

We specified how we determined our sample size, all data exclusions, all manipulations, and all measures in the associated preregistration (https://doi.org/10.17605/OSF.IO/76PV5), published on June 9, 2023, thus before we started data collection on June 12, 2023. Please note that the numbering and wording of hypotheses was slightly modified from the preregistration to increase clarity, while not affecting their content. Preprocessed data, analysis scripts and supplemental materials can be found in the associated OSF repository (https://osf.io/ascqx/).

2.4 Results

2.4.1 Demography, musicality, and personality of participants

First, we checked for important demographic and psychological variables that the two groups were comparable. Singers and instrumentalists did not differ significantly in the socioeconomic status assessed via educational level, X^2 (2, N = 88) = 1.06, p = .588; highest academic degree, X^2 (7, N = 88) = 9.06, p = .249, and household income, X^2 (4, N = 88) = 5.23, p = .264 (for more details see Table S2 on OSF). Further, the groups did not differ in age or positive and negative affect (assessed with the PANAS) and were comparable regarding Big Five personality traits and autistic traits. In the Gold-MSI, singers and instrumentalists scored comparatively on the general musicality score, but there were differences on two subfactors: instrumentalists scored higher on the subfactor Formal Education, while singers scored higher on Singing. In the PROMS, both groups performed comparably in all four subtests. Participant characteristics assessed via self-report and music performance in the PROMS are summarized in **Table 1**.

Table 1

Characteristics of participants - Demography, personality, and musicality

	Singers	**Instrumentalists**				
	M (SD)	M (SD)	t	df[a]	p	Cohen‘s d
Age	27.02 (8.2)	28.51 (10.6)	-0.73	78.93	.465	-0.17 [-0.61, 0.28]
PANAS						

positive Affect	3.00 (0.68)	3.11 (0.57)	-0.77	84.78	.446	-0.17 [-0.59, 0.26]	
negative Affect	1.53 (0.47)	1.40 (0.35)	1.49	80.61	.141	0.33 [-0.11, 0.77]	
Big Five							
Openness	4.04 (0.55)	3.99 (0.51)	0.46	85.96	.647	0.10 [-0.32, 0.52]	
Conscientiousness	3.47 (0.69)	3.76 (0.70)	-1.91	85.62	.060	-0.41 [-0.84, 0.02]	
Extraversion	3.21 (0.70)	3.00 (0.73)	1.44	85.3	.155	0.31 [-0.12, 0.74]	
Agreeableness	3.81 (0.57)	4.01 (0.60)	-1.61	85.29	.112	-0.35 [-0.77, 0.08]	
Neuroticism	2.74 (0.77)	2.61 (0.78)	0.80	85.75	.426	0.17 [-0.25, 0.60]	
AQ							
Total	18.2 (6.15)	19.28 (8.55)	-0.68	76.04	.500	-0.16 [-0.60, 0.30]	
Attention to Detail	5.4 (2.33)	5.63 (2.53)	-0.44	84.64	.662	-0.10 [-0.52, 0.33]	
Social	12.8 (5.37)	13.65 (7.53)	-0.61	75.65	.545	-0.14 [-0.59, 0.31]	
Social Skills	2.4 (1.94)	3.09 (2.95)	-1.30	72.01	.200	-0.31 [-0.77, 0.16]	
Communication	2.53 (1.94)	2.44 (2.31)	0.20	82.02	.842	0.04 [-0.39, 0.48]	
Imagination	2.51 (1.75)	2.81 (1.88)	-0.78	84.86	.437	-0.17 [-0.60, 0.26]	
Attention Switching	5.36 (1.91)	5.30 (2.23)	0.12	82.69	.905	0.03 [-0.40, 0.46]	
Gold-MSI							
General ME	4.78 (0.85)	4.75 (0.80)	0.17	85.99	.866	0.04 [-0.39, 0.46]	
Active Engagement	3.83 (0.82)	4.21 (1.13)	-1.79	76.79	.078	-0.41 [-0.86, 0.05]	
Formal Education	**4.39 (1.14)**	**4.95 (0.62)**	-2.85	68.31	.006	-0.69 [-1.18, -0.20]	**
Emotion	5.50 (0.81)	5.60 (0.76)	-0.60	85.99	.549	-0.13 [-0.55, 0.29]	
Singing	**4.98 (0.97)**	**4.19 (1.27)**	3.25	78.56	.002	0.73 [0.27, 1.19]	**
Perception	5.73 (0.82)	5.77 (1.03)	-0.22	80.11	.825	-0.05 [-0.49, 0.39]	
PROMS							
Pitch	0.23 (0.08)	0.24 (0.06)	-0.30	82.92	.766	-0.07 [-0.50, 0.37]	
Melody	0.17 (0.10)	0.14 (0.10)	1.29	85.08	.199	0.28 [-0.15, 0.71]	
Timbre	0.29 (0.08)	0.3 (0.09)	-0.59	86.00	.556	-0.13 [-0.55, 0.30]	
Rhythm	0.31 (0.09)	0.32 (0.09)	-0.56	85.13	.577	-0.12 [-0.55, 0.30]	

Note. Descriptive values show mean ratings for the PANAS (Breyer & Bluemke, 2016), the Big-Five Domains (Rammstedt et al., 2018), and the Gold-MSI (Müllensiefen et al., 2014). AQ score were calculated based on Hoekstra et al. (2008) and Baron-Cohen et al. (2001).

[a] *Note that original degrees of freedom were 86 but were corrected due to unequal variance.*

2.4.2 Emotion classification performance

The mean proportion of correct responses was submitted to an ANOVA with Emotion (Happiness, Pleasure, Fear, and Sadness) and Morph Type (Full, F0, and Timbre) as repeated measures factors and Group (singers and instrumentalists) as a between subject factor (see **Table 2**).

The results revealed main effects of **Emotion** and **Morph Type**, which were qualified by an interaction. Crucially, however, we found no main effects or interactions involving **Group** (see **Figure 2**), which was also confirmed by a Bayesian ANOVA (see Table S7 on OSF). Planned Bayesian analysis revealed moderate evidence for the null effect of group for overall performance (p = .542, BF_{10} = 0.265), as well as for Full (p = .392, BF_{10} = 0.310), F0 (p = .935, BF_{10} = 0.226), and Timbre morphs (p = .555, BF_{10} = 0.262) separately. Thus, we found evidence consistent with our hypotheses H1 and H2.

Table 2

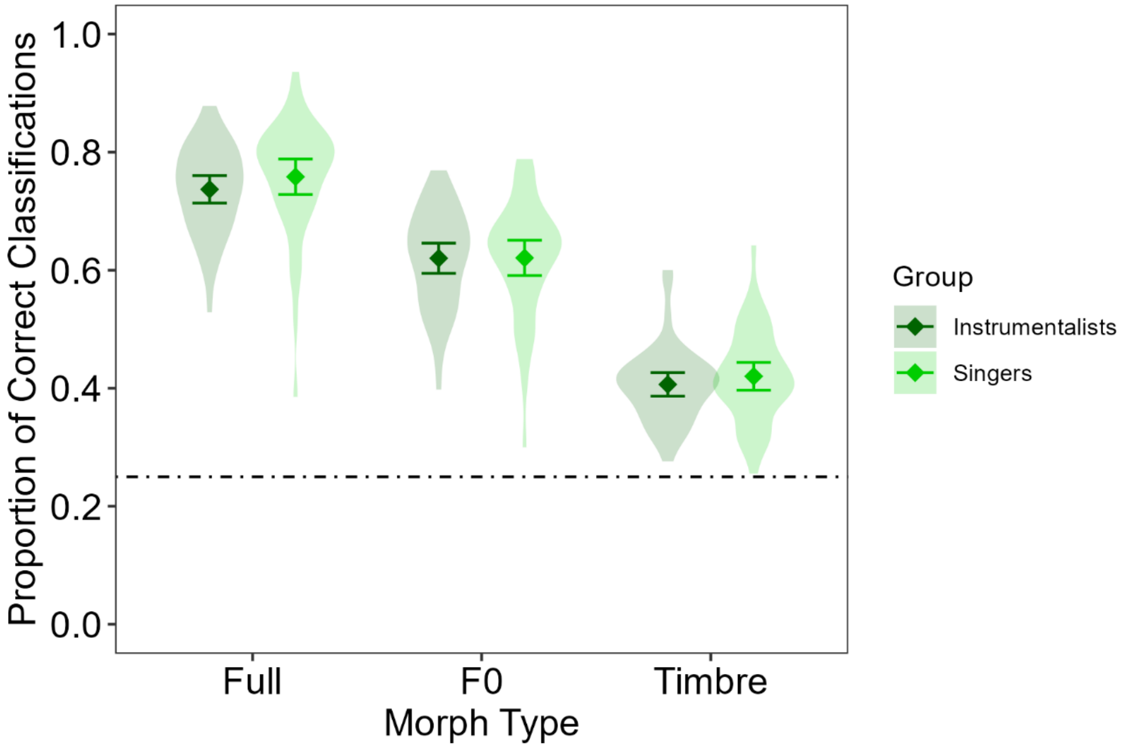
Results of the 4 × 3 × 2 mixed-effects ANOVA on the mean proportion of correct responses

	df1	df2	F	p	Ω_p^2 [95%-CI]	ε_{HF}
Group	1	86	0.38	.542	.00 [.00, .01]	
Emotion	3	258	72.43	<.001	.45 [.36, .52]	
Morph Type	2	172	768.93	<.001	.90 [.87, .92]	.741
Group x Emotion	3	258	2.14	.095	.01 [.00, .04]	
Group x Morph Type	2	172	0.36	.635	.00 [.00, .01]	
Emotion x Morph Type	6	516	22.78	<.001	.20 [.14 .25]	.827
Group x Emotion x Morph Type	6	516	1.33	.249	.00 [.00 .01]	

Note. *ε_{HF} = Huynh–Feldt (HF) epsilon correction factor in case of violation of the sphericity assumption.*

Figure 2

Mean proportion of correct responses per Morph Type separately for singers and instrumentalists

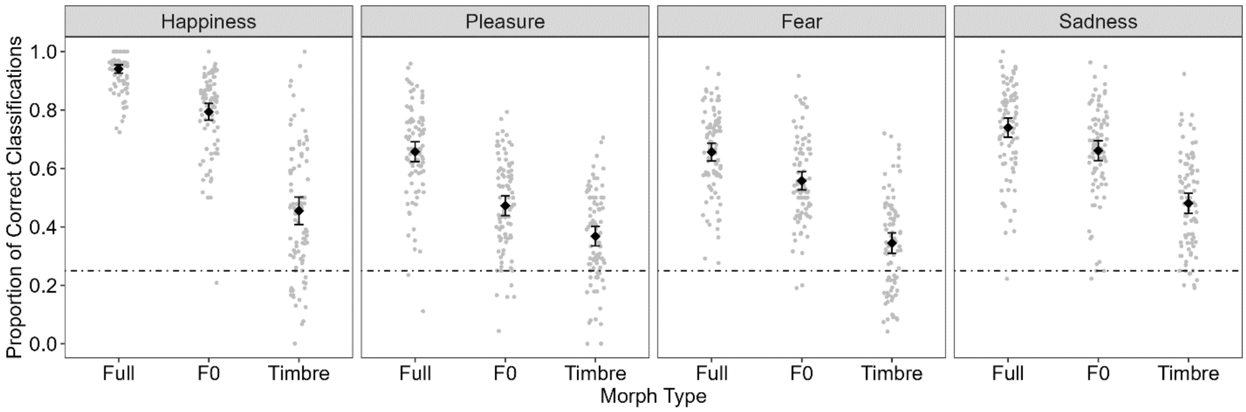


Note. Whiskers represent 95% confidence intervals. Violin plots represent variation of individual participants. The dotted line represents guessing rate at .25.

Follow-up analysis of the Morph Type effect revealed that performance was best in the Full condition (*M* = 0.75 ± 0.01 SEM), followed by the F0 (*M* = 0.62 ± 0.01) and then the Timbre condition (*M* = 0.41 ± 0.01); Full vs. F0: $|t(87)| = 23.28$, $p < .001$, $d = 2.50$ [2.07, 2.92], F0 vs Timbre: $|t(87)| = 21.44$, $p < .001$, $d = 2.30$ [1.90, 2.70], Full vs Timbre: $|t(87)| = 33.94$, $p < .001$, $d = 3.64$ [3.06, 4.21]). This Morph Type main effect was also found for all emotions separately (all Fs(2, 174) > 102.44, $p < .001$), although it differed slightly between emotions, as suggested by the interaction (see **Figure 3,** for all post-hoc tests, refer to the full analysis on OSF). Performance difference between F0 and Timbre was largest for Happiness ($M_{F0\text{-}Timbre}$ = 0.34 ± 0.02 SEM), followed by Fear ($M_{F0\text{-}Timbre}$ = 0.21 ± 0.02), Sadness ($M_{F0\text{-}Timbre}$ = 0.18 ± 0.02), and Pleasure ($M_{F0\text{-}Timbre}$ = 0.10 ± 0.02; all pairwise comparisons $|ts(77)| \geq 2.57$, $ps \leq .012$, $ds \geq 0.28$ [0.06 0.49], except for Fear vs. Sadness ($|t(87)| = 1.13$, p = .261). These effects of Morph Type and Emotion therefore present a full replication of the patterns reported in Nussbaum et al. (2024). For confusion matrices and supplemental analyses, please refer to Figures S1-S3 on OSF.

Figure 3

Mean proportion of correct responses per Emotion and Morph Type



Note. Whiskers represent 95%-confidence intervals. Grey dots represent individual participants' data. The dotted line represents guessing rate at .25.

3 Part II: Correlational analyses

3.1 Hypotheses

In Part II, we focused on the correlations between auditory sensitivity and vocal emotion recognition. The aim here was to see if the patterns found in Nussbaum et al. (2024) would replicate. Therefore, we formulated the following predictions:

Correlations with the PROMS

H3a: Averaged vocal emotion recognition (VER) performance is correlated with averaged music perception performance.

H3b: Full-VER and F0-VER are correlated with melody perception in music

Correlations with the GOLD-MSI:

H3c: Averaged-VER and Full-VER are correlated with the general musical sophistication index (General-ME), measured by the Gold-MSI.

H3d: Averaged-VER and Full-VER are correlated with the perception subscale of the Gold-MSI

H3e: Averaged-VER and Full-VER are correlated with self-rated singing abilities subscale of the Gold-MSI.

3.2 Data analysis

We calculated Spearman correlations between vocal emotion recognition performance and both the PROMS music perception performance and the Gold-MSI self-rated musicality. P-values were adjusted for multiple comparisons using the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995). Following our pre-registered plan, all correlations were controlled for formal musical education as well, but we found that this made no difference to the observed patterns.

3.3 Results

Replicating our previous findings, we obtained a strong correlation between vocal emotion recognition and music perception performance, as measured with the PROMS (Table 3, and Table S10 on OSF). Further, we replicated the specific link between vocal emotion recognition and the melody subtest. Although we did not have a specific hypothesis, we again found a link with Rhythm perception. There were no links between performance in the timbre morph condition and the timbre subtest. Overall, this pattern of correlations almost completely replicates the pattern observed in a previous sample (Nussbaum et al., 2024). Thus, the link between auditory sensitivity and vocal emotion recognition seems to be highly comparable across professional musicians, non-musicians and amateurs.

Table 3

Spearman correlations between the PROMS and vocal emotion recognition performance

	$PROMS_{Avg}$	Pitch	Melody	Timbre	Rhythm
VER_{Avg}	**.39 (.002)**	.17 (.142)	**.29 (.017)**	.23 (.050)	**.38 (.002)**
Full-Morphs	**.34 (.005)**	.14 (.219)	**.27 (.022)**	.23 (.050)	**.31 (.009)**
F0-Morphs	**.39 (.002)**	.16 (.186)	**.34 (.005)**	**.24 (.044)**	**.32 (.008)**
Timbre-Morphs	.22 (.054)	.12 (.305)	.08 (.473)	.10 (.352)	**.25 (.039)**

Note. VER = Vocal Emotion Recognition performance. p-values were adjusted for multiple comparisons using the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995). VER_{Avg}: VER performance averaged across all trials, Full-Morphs: VER in the Full Morph condition only, F0-Morphs: VER in the F0 Morph condition only, Timbre-Morphs: VER in the Timbre Morph condition only, $PROMS_{Avg}$: music perception performance averaged across all four subtests of the PROMS (Pitch, Melody, Timbre, and Rhythm).

In contrast, we found no links between vocal emotion recognition performance and self-rated musicality, as measured by the Gold-MSI (all correlations ≤ 0.21, *p*s ≥ .051, both with and without correction for formal musical education, details on Tables S11 and S12 on OSF). Thus, for amateurs, we could not replicate the link with self-rated musical sophistication, perception and singing abilities, which we observed in our previous sample of professional musicians and non-musicians. Therefore, we found evidence for our hypotheses H3a and H3b, but not for hypotheses H3c – H3e.

4 Part III: Comparison of professionals, amateurs and non-musicians

In the third part, we compared amateur musicians to both professional musicians and non-musicians. Mostly, we predicted that amateurs and professional musicians would be comparable regarding vocal emotion recognition. However, as previous evidence reviewed above showed that amateurs can differ from professionals in cognitive abilities which could be linked to emotional sensitivity, we also considered the option that amateurs could be more proficient at making emotional inferences than professionals, reflected in H5 below. Compared to our group of non-musicians (H4), we assumed that amateurs would outperform them when emotions were expressed via full emotion cues and F0 cues only, but not timbre, because this is exactly the pattern we observed for the comparison of professional musicians and non-musicians in Nussbaum et al. (2024).

4.1 Hypotheses

H4: Amateur musicians outperform non-musicians in vocal emotion recognition, in the Full and in the F0 condition.

H5: Amateurs perform equal or better to professional musicians in the Full and the F0 condition.

4.2 Method

For this analysis, we collapsed all participants from Part I into the group of amateur musicians and compared it to the groups of professional musicians and non-musicians reported in Nussbaum et al. (2024). Note that we added one participant to the professional group, because he held a master's degree in music (see Part I), so numbers slightly diverge from the original publication. All professional musicians reported to have a music-related academic degree or a non-academic music qualification. Non-musicians played no instrument or engaged in any other musical activities. For a more detailed description, please refer to Nussbaum et al. (2024).

In total, we analyzed data from 40 professional musicians (20 male, 20 female, aged 20 to 42 years [*M* = 29.6; *SD* = 5.58]), 38 non-musicians (18 male, 20 female, aged 19 to 48 years [*M* = 30.5; *SD* = 6.54]) and 88 amateurs (40 male, 46 female, 2 diverse, aged 18 to 54 years [*M* = 27.8; *SD* = 9.44]).

The stimulus material, design and data analysis were identical to Part I. We focused our analysis on the comparison of amateurs with the other two groups, because the comparison of professional musicians and non-musicians is reported in Nussbaum et al. (2024).

4.3 Results

4.3.1 Demography, musicality, and personality of participants

Again, we first confirmed that the groups were comparable on important individual variables. Professionals, amateurs and non-musicians did not differ in the socioeconomic status assessed via educational level ($\chi^2(6, N = 166) = 11.11$, $p = .085$) and highest academic degree ($\chi^2(16, N = 166) = 24.04$, $p = .089$). However, there were differences regarding household income ($\chi^2(8, N = 166) = 20.19$, $p = .010$, Cramer's V = .25), with amateurs reporting higher household income than professionals and non-musicians.

Participant characteristics are summarized in **Table 4.** For a full report of statistical details, please refer to Tables S13-S21 on OSF. The groups were comparable in age as well as in positive and negative affect (assessed with the PANAS). For the Big Five, analyses of variance revealed group differences for extraversion, with slightly higher levels in professionals than in amateurs. Regarding autistic traits, the three groups did not differ in their overall score, but there were differences on several subscales. In the Gold-MSI, professional musicians scored significantly higher than amateurs on all subscales, which in turn scored higher than non-musicians. This is a pattern (professionals > amateurs > non-musicians) one would expect for self-rated musicality. In the PROMS, professionals outperformed amateurs in the Pitch and Melody subtest, whereas there were no differences in the Timbre and Rhythm subtests. Amateurs performed better than non-musicians in the Pitch, Melody and Rhythm subtest but not in the Timbre subtest. Thus, a clear pattern of professionals > amateurs > non-musicians was only found for melody and pitch.

Table 4

Characteristics of participants - Demography, personality, and musicality

	Professionals		Amateurs		Non-Musicians
	M (SD)		M (SD)		M (SD)
PANAS					
positive Affect	3.32 (0.65)		3.05 (0.63)		3.1 (0.67)
negative Affect	1.69 (0.48)		1.47 (0.42)		1.49 (0.69)
Big Five					
Openness	4.12 (0.50)		4.02 (0.53)		3.81 (0.80)
Conscientiousness	3.49 (0.71)		3.61 (0.70)		3.76 (0.72)
Extraversion	**3.48 (0.66)**	>	**3.11 (0.72)**		3.38 (0.79)
Agreeableness	3.92 (0.57)		3.91 (0.59)		3.75 (0.66)
Neuroticism	2.95 (0.65)		2.69 (0.77)		2.58 (0.82)
AQ					
Total	15.7 (4.98)		18.73 (7.40)		17.58 (6.41)
Attention to Detail	5.43 (2.04)		**5.51 (2.42)**	>	**4.32 (2.01)**
Social	**10.28 (4.70)**	<	**13.22 (6.49)**		13.26 (6.51)
Social Skills	**1.48 (1.68)**	<	**2.74 (2.49)**		2.61 (2.63)
Communication	1.85 (1.61)		2.49 (2.12)		2.39 (1.73)
Imagination	2.18 (1.52)		2.66 (1.81)		2.87 (1.95)
Attention Switching	4.78 (1.91)		5.33 (2.06)		5.39 (1.92)
Gold-MSI					
General ME	**5.68 (0.50)**	>	**4.76 (0.82)**	>	**2.74 (1.07)**
Active Engagement	**4.94 (0.81)**	>	**4.02 (1.00)**	>	**2.95 (1.19)**
Formal Education	**5.95 (0.56)**	>	**4.66 (0.96)**	>	**1.71 (0.68)**
Emotion	**5.88 (0.73)**	>	**5.55 (0.78)**	>	**4.95 (1.32)**
Singing	**5.34 (0.83)**	>	**4.59 (1.19)**	>	**2.84 (1.26)**
Perception	**6.31 (0.51)**	>	**5.75 (0.92)**	>	**4.22 (1.49)**
PROMS					
Pitch	**0.27 (0.06)**	>	**0.24 (0.07)**	>	**0.18 (0.06)**
Melody	**0.23 (0.08)**	>	**0.16 (0.10)**	>	**0.07 (0.08)**
Timbre	0.32 (0.08)		0.29 (0.08)		0.26 (0.09)
Rhythm	0.33 (0.08)		**0.32 (0.09)**	>	**0.27 (0.08)**

Note. Descriptive values show mean ratings for the PANAS (Breyer & Bluemke, 2016), the Big-Five Domains (Rammstedt et al., 2018), and the Gold-MSI (Müllensiefen et al., 2014). AQ score were calculated based on Hoekstra et al. (2008) and Baron-Cohen et al. (2001).

Comparison signs (">" or "<") indicate significant differences. For a full report of statistical details, please refer to OSF. For a detailed description and discussion of the differences between professional musicians and non-musicians, please refer to the previous publication (Nussbaum et al., 2024).

4.3.2 Emotion classification performance

The mean proportion of correct responses was submitted to an ANOVA with Emotion (Happiness, Pleasure, Fear, and Sadness) and Morph Type (Full, F0, and Timbre) as repeated measures factors and Group (professionals, amateurs, and non-musicians) as a between subjects-factor (see **Table 5**). For the parallel Bayesian ANOVA, please refer to S8 on OSF. The results revealed significant main effects of **Emotion** and **Morph Type**, which were qualified by an interaction between **Emotion** and **Morph Type**. Please note that these effects were already present in the two datasets that entered the current data (reported in Nussbaum et al., 2024 and Part I above). Because they afford no new information, they are not further detailed here. Crucially, however, we found no main effect involving **Group** and only a trend for an interaction between **Group and Morph Type**.

Table 5

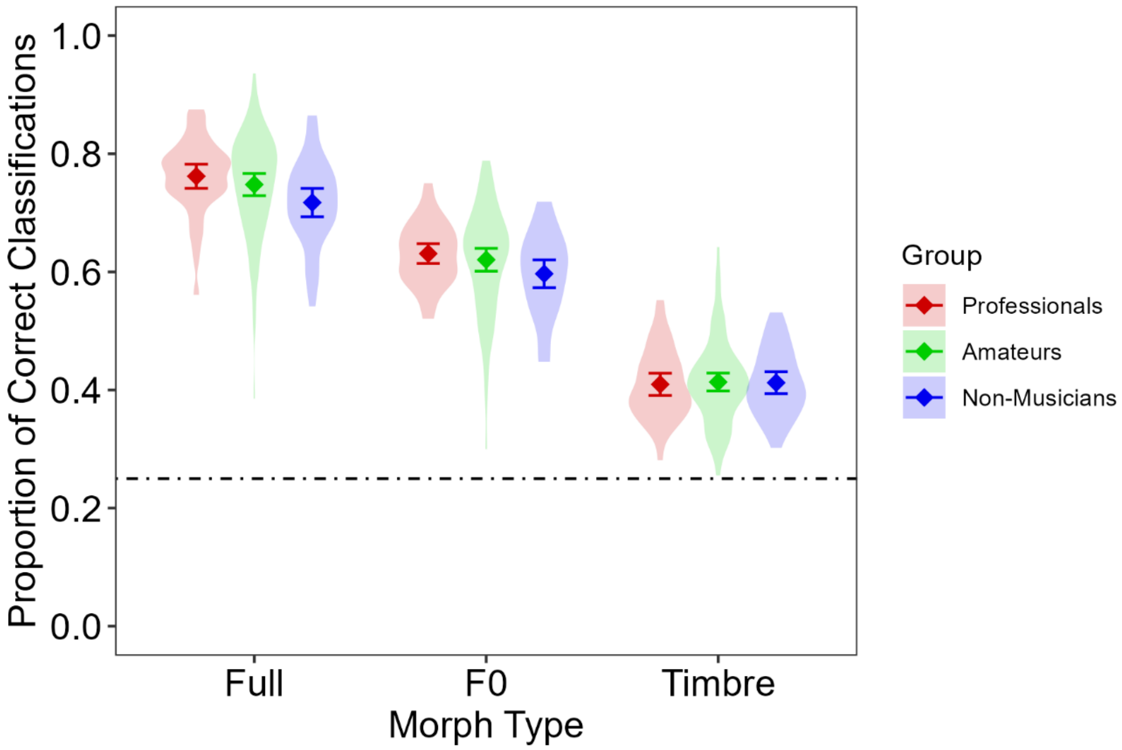
Results of the 4 × 3 × 3 mixed-effects ANOVA on the mean proportion of correct responses

	df1	df2	F	p	Ω_p^2 [95%-CI]	ε_{HF}
Group	2	163	1.96	.144	.01 [.00, .06]	
Emotion	3	489	130.24	<.001	.44 [.38, .49]	
Morph Type	2	326	1357.80	<.001	.89 [.87, .91]	.829
Group x Emotion	6	489	1.17	.322	.00 [.00, .01]	
Group x Morph Type	4	326	2.14	.089	.01 [.00, .04]	.829
Emotion x Morph Type	6	978	40.95	<.001	.20 [.15, .24]	.875
Group x Emotion x Morph Type	12	978	0.74	.688	.00 [.00, .01]	.865

Planned comparisons between **professionals and amateurs** revealed moderate evidence for the null effect of overall performance (p = .473, BF_{10} = 0.238), as well as for Full (p = .322, BF_{10} = 0.286), F0 (p = .435, BF_{10} = .245), and Timbre morphs (p = .840, BF_{10} = 0.205) separately. Planned comparisons between amateurs and non-musicians revealed inconclusive evidence for overall performance (p = .107, BF_{10} = 0.546), as well as for Full (p = .044, BF_{10} = 1.009), and F0 (p = .105, BF_{10} = 0.576) separately. For Timbre morphs, there was moderate evidence for the null effect (p = .975, BF_{10} = 0.205). Thus, we found evidence consistent with H5, but inconclusive evidence regarding H4.

Figure 4

Mean proportion of correct responses per Morph Type separately for professionals, amateurs, and non-musicians



Note. Whiskers represent 95% confidence intervals. Violin plots represent variation of individual participants. The dotted line represents guessing rate at .25.

5 Discussion

In the present study, we shed new light on the link between musicality and vocal emotion recognition, by focusing on different subgroups of musicians. In line with our hypotheses, emotion recognition was found to be comparable between both singers and instrumentalists, and also between amateur musicians and professionals. We further replicated the consistent link between sensitivity towards musical patterns and vocal emotion recognition, although this was reflected in objective performance measures only, but not in self-rated musicality. In total, these results suggest that the link between musicality and vocal emotion recognition is associated with individual differences in auditory sensitivity, which is not tied to a particular type or amount of musical activity. In what follows, we will discuss these findings in more detail.

5.1 Singers vs. instrumentalists

For the present study, we recruited a well-powered sample of singers and instrumentalists, which were comparable with regard to personality, socio-economic background, and objective music perception performance. They only differed in two aspects of self-rated musicality: unsurprisingly, singers scored higher on self-rated singing abilities, and instrumentalists scored higher on self-rated formal education. The latter may be because mastering an instrument to the point where one can play in an ensemble arguably takes more formal training than being able to sing in a choir.

In line with our prediction, the vocal emotion recognition performance of singers and instrumentalists did not differ in any condition, as supported by Bayesian analysis with moderate evidence. This finding does not simply represent a general absence of effects. In fact, we replicated the strong and consistent effect of vocal parameters (manipulated via voice morphing) on emotion recognition, with F0 contour being relatively more informative than timbre, across all emotions, while performance was still best in the Full condition. The fact that this pattern was found to be highly comparable for singers and instrumentalists suggests that they use the same acoustic cues to perceive vocal emotions and do so with comparable efficiency. Note that the effects of the F0 and timbre manipulation on emotion recognition also provided a complete replication of the pattern observed in professional musicians and non-musicians in our previous study (Nussbaum et al., 2024). For an even more detailed reflection on the roles of F0 and timbre for emotion recognition, both on a behavioral and neural level (irrespective of musicality), please also refer to Nussbaum et al. (2022).

Although this is precisely what we predicted, we want to address two observations from the literature which may seem incompatible with our findings: First, our findings diverge from Thompson et al. (2004), who found that singing lessons may interfere with vocal emotional processing. However, this study had several methodological limitations including a very small sample size and large drop-out rates. With the present, substantially powered design, we do not see any evidence for a disruptive effect of active singing. This aligns with the broader literature on protective effects of singing on cognitive and socio-emotional functioning (Moisseinen et al., 2024; Tragantzopoulou & Giannouli, 2025). Second, several studies found correlations between singing abilities and emotion recognition (Correia et al., 2022; Greenspon & Montanaro, 2023; Nussbaum et al., 2024). However, all these studies included both singers and instrumentalists. In fact, it is quite common that instrumentalists have good singing abilities as well, although it is not their preferred form of musical expression. Thus, while we do not argue against a relationship between singing abilities and emotion recognition per se, we lack evidence that emotion recognition is more pronounced in specific groups of musicians, or that this is causally linked to certain musical activities. Instead, this correlation seems to be mediated via natural auditory sensitivity. Note that we failed to find the predicted correlation between self-rated singing abilities and emotion recognition in the present study. This may be attributed to potentially decreased variance in our group of amateurs with regard to self-rated musicality (Gold-MSI). In fact, we did not find any correlations with the Gold-MSI subscales in amateurs, which is in contrast to the pattern we had observed in the sample comprised of professionals and non-musicians (Nussbaum et al., 2024).

5.2 Amateurs compared to professional musicians and non-musicians

As predicted, we found evidence that emotion perception performance does not differ between amateur musicians and professionals, further strengthening the notion that the amount of music training is not a major influence on vocal emotion recognition abilities (Schellenberg & Lima, 2024). We also had hypothesized that amateurs would outperform non-musicians, because professionals outperformed them in our previous study (Nussbaum et al., 2024). This comparison yielded inconclusive evidence. We speculate that the present amateur sample was more heterogenous than the professional one, and as a result, our design may have lacked statistical power to detect potentially small differences between amateurs and non-musicians, despite the substantial sample size. This issue may be resolved in a follow-up study.

In addition to the direct group comparison, we performed a correlational analysis on our sample of amateurs (Part II) which was in parallel to an analysis we had performed previously on professionals and non-musicians (Nussbaum et al., 2024). Importantly, we found a highly similar pattern of correlations between vocal emotion recognition and music perception abilities, especially for melody and rhythm. We therefore conclude that individual differences in music perception abilities play an important role, irrespective of the assignment to any (non)-musical group. This is fully in line with the current literature, which consistently emphasizes the role of auditory sensitivity and argues against a causal effect of training or musical activity on vocal emotion recognition (Neves et al., 2025; Schellenberg & Lima, 2024).

5.3 Directions for future research

To the best of our knowledge, this study is the first substantially powered comparison between different subgroups of musicians (specifically, singers vs. instrumentalists and professionals vs. amateurs) with respect to vocal emotion recognition and therefore closes an open gap in the literature. Nevertheless, the present study has several limitations which deserve consideration. First, our dataset was limited to four emotions expressed through short pseudowords. Future research should examine the extent to which these findings generalize to other types of vocal material. Second, all musicians were socialized in Western music culture and fluent German speakers, therefore findings may not generalize to other music and language backgrounds (Morrison & Demorest, 2009). Note, however, that links between music perception performance and emotion recognition irrespective of formal musical training have been observed in a substantial Portuguese sample of musically trained and untrained participants who varied widely in their musical skill (Correia et al., 2022).

What posed a particular challenge in the present study was the recruitment of mutually exclusive groups of musicians. While the distinction between singers and instrumentalists may seem trivial at first glance, it is actually rare that musicians engage in one of these activities only. Thus, for practical reasons, we recruited participants who self-assigned clearly to one form and made sure that their current activity is either one or the other. Yet, some of the singers reported that they used to play an instrument as well. In a similar vein, the distinction between amateurs and professionals is not fully straightforward and may represent a continuum rather than clear-cut categories. Some individuals pursue music as a profession, but without a formal music degree, others vice versa (Zendel & Alexander, 2020). Some people transition from amateurs to a professional level later in life (Taylor & Hallam, 2011). Thus, while the argument still holds that future research should consider the heterogeneity of musicians, it may be more adequate to pursue variability on a spectrum rather than for distinct groups.

On a practical note, we must acknowledge the technical randomization error. While in the previous study (Nussbaum et al., 2024) stimuli were drawn only once, as intended, the present code allowed full randomization with duplication and omissions of stimuli. While undoubtedly unfortunate, we are nevertheless confident that this error has not affected our results substantially. First, the classification patterns for different Morph Types and Emotions fully replicate our previous study (cf. Figure 3) and we observed highly similar correlations between vocal emotion recognition and music perception performance (cf. Table 3). Second, while this issue might have decreased our signal-to-noise ratio, it would not have introduced a specific bias. Thus, we still consider both studies sufficiently comparable.

While the present behavioral data did not reveal differences between musical subgroups, future research may unravel more fine-grained patterns in the brain. Several studies have reported brain differences in auditory and motor processing between amateurs and professionals (Kleber et al., 2010; Lotze et al., 2003; Oechslin et al., 2013; Papadaki et al., 2023) and singers and instrumentalists (Halwani et al., 2011), but as yet such studies do not provide insights into vocal emotional processing. While we observed differences in electrophysiological responses in professional musicians and non-musicians during vocal emotion recognition (Lehnen et al., 2025; Nussbaum et al., 2023), a meaningful fine-grained comparison between singers and instrumentalists, or amateurs and professionals, would arguably require a much bigger sample.

Finally, with respect to the individual dynamics of the tight neural overlap of expression and perception in vocal communication, we hold that it may be worthwhile to focus on another group of vocal experts: professional voice actors or imitators. This is because such experts, unlike singers, are specifically trained to express emotions in the *spoken* voice. In fact, a recent study found that voice actors show enhanced sensitivity for linguistic voice prosody and enhanced neural tracking for voice pitch compared to non-actors (Kachlicka & Tierney, 2024). Vocal acting usually involves deliberate exaggeration of vocal expressions and may therefore require a fine-grained explicit representation of vocal emotions (Berry & Brown, 2019). While there is research comparing actors and non-actors regarding the expression of emotions (Jürgens et al., 2015), we do not know how this is mirrored in vocal emotion recognition performance and how this group may differ from individuals with singing expertise.

6 Summary and Conclusion

In the present study, we investigated how emotion perception differs in musical subgroups and compared singers and instrumentalists as well as amateurs and professionals. In line with our prediction, we observed no differences between musical subgroups, suggesting that emotion perception abilities are not primarily tied to the type or amount of musical activity. Instead, we replicated the link between vocal emotion recognition and music perception abilities, especially for melodies. This adds a new perspective to the accumulating evidence that the link between musicality and vocal emotion recognition is predominantly associated with individual differences in natural auditory sensitivity.

7 Acknowledgements

This study was conducted by J.D. in partial fulfilment of the requirements for a master's thesis. The original voice recordings that served as a basis for creating our stimulus material were provided by Sascha Frühholz. We thank Hannah Strauß for support with the PROMS. We are grateful to all participants of the study.

8 Conflicts of Interests

The authors declare no competing interests.

9 Declaration of generative AI and AI-assisted technologies in the writing process

No generative AI or AI-assisted technologies were used in the writing process of this manuscript.

10 Credit Author Statement

Christine Nussbaum – Conceptualization, Methodology, Software, Visualization, Formal analysis, Writing - Original Draft, Supervision

Jessica Dethloff - Data collection, Formal analysis, Visualization, Writing - Original Draft

Annett Schirmer – Methodology, Writing - Review & Editing, Supervision

Stefan R. Schweinberger – Conceptualization, Writing - Review & Editing, Supervision

11 Resource availability

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Christine Nussbaum (christine.nussbaum@uni-jena.de).

Data: All raw and preprocessed data have been deposited at the associated OSF repository (https://osf.io/ascqx/) and are publicly available.

Code: All original code for data analysis has been deposited at the associated OSF repository (https://osf.io/ascqx/) and is publicly available.

Additional information: Supplemental figures and tables have been deposited at the associated OSF repository (https://osf.io/ascqx/) and are publicly available.

References

Akkermans, J., Schapiro, R., Müllensiefen, D., Jakubowski, K., Shanahan, D., Baker, D., Busch, V., Lothwesen, K., Elvers, P., Fischinger, T., Schlemmer, K., & Frieler, K. (2019). Decoding emotions in expressive music performances: A multi-lab replication and extension study. *Cognition & Emotion*, *33*(6), 1099–1118. https://doi.org/10.1080/02699931.2018.1541312

Baldé, A. M., Lima, C. F., & Schellenberg, E. G. (2025). Associations between musical expertise and auditory processing. *Journal of Experimental Psychology: Human Perception and Performance*, *51*(6), 747–763. https://doi.org/10.1037/xhp0001312

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614–636. https://doi.org/10.1037/0022-3514.70.3.614

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J.-C., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Berry, M., & Brown, S. (2019). Acting in action: Prosodic analysis of character portrayal during acting. *Journal of Experimental Psychology. General*, *148*(8), 1407–1425. https://doi.org/10.1037/xge0000624

Boersma, P. (2018). Praat: Doing phonetics by computer [Computer program]: Version 6.0.46, retrieved January 2020 from http://www.praat.org/. *Http://www. Praat. Org*.

Bonde, L. O., Juel, K., & Ekholm, O. (2018). Associations between music and health-related outcomes in adult non-musicians, amateur musicians and professional musicians—Results from a nationwide Danish study. *Nordic Journal of Music Therapy*, *27*(4), 262–282. https://doi.org/10.1080/08098131.2018.1439086

Breyer, B., & Bluemke, M. (2016). *Deutsche Version der Positive and Negative Affect Schedule PANAS (GESIS Panel).* https://doi.org/10.6102/zis242

Christiner, M., & Reiterer, S. M. (2015). A Mozart is not a Pavarotti: Singers outperform instrumentalists on foreign accent imitation. *Frontiers in Human Neuroscience*, *9*, 482. https://doi.org/10.3389/fnhum.2015.00482

Correia, A. I., Castro, S. L., MacGregor, C., Müllensiefen, D., Schellenberg, E. G., & Lima, C. F. (2022). Enhanced recognition of vocal emotions in individuals with naturally good musical abilities. *Emotion*, *22*(5), 894–906. https://doi.org/10.1037/emo0000770

Fisher, R. A., Hoult, A. R., & Tucker, W. S. (2020). A Comparison of Facial Muscle Activation for Vocalists and Instrumentalists. *Journal of Music Teacher Education*, *30*(1), 53–64. https://doi.org/10.1177/1057083720947412

Freitag, C. M., Retz-Junginger, P., Retz, W., Seitz, C., Palmason, H., Meyer, J., Rösler, M., & Gontard, A. von (2007). Evaluation der deutschen Version des Autismus-Spektrum-Quotienten (AQ) - die Kurzversion AQ-k. *Zeitschrift Für Klinische Psychologie Und Psychotherapie*, *36*(4), 280–289. https://doi.org/10.1026/1616-3443.36.4.280

Frühholz, S., & Schweinberger, S. R. (2021). Nonverbal auditory communication - Evidence for integrated neural systems for voice signal production and perception. *Progress in Neurobiology*, *199*, 101948. https://doi.org/10.1016/j.pneurobio.2020.101948

Greenspon, E. B., & Montanaro, V. (2023). Singing ability is related to vocal emotion recognition: Evidence for shared sensorimotor processing across speech and music. *Attention, Perception & Psychophysics*, *85*(1), 234–243. https://doi.org/10.3758/s13414-022-02613-0

Hake, R., Kreutz, G., Frischen, U., Schlender, M., Rois-Merz, E., Meis, M., Wagener, K. C., & Siedenburg, K. (2024). A Survey on Hearing Health of Musicians in Professional and Amateur Orchestras. *Trends in Hearing*, *28*, 23312165241293762. https://doi.org/10.1177/23312165241293762

Halwani, G. F., Loui, P., Rüber, T., & Schlaug, G. (2011). Effects of practice and experience on the arcuate fasciculus: Comparing singers, instrumentalists, and non-musicians. *Frontiers in Psychology*, *2*, 156. https://doi.org/10.3389/fpsyg.2011.00156

Hoekstra, R. A., Bartels, M., Cath, D. C., & Boomsma, D. I. (2008). Factor structure, reliability and criterion validity of the Autism-Spectrum Quotient (AQ): A study in Dutch population and patient groups. *Journal of Autism and Developmental Disorders*, *38*(8), 1555–1566. https://doi.org/10.1007/s10803-008-0538-x

Jarosz, A. F., & Wiley, J. (2014). What Are the Odds? A Practical Guide to Computing and Reporting Bayes Factors. *The Journal of Problem Solving*, *7*(1). https://doi.org/10.7771/1932-6246.1167

JASP Team. (2025). *JASP (Version 0.19.3)[Computer software]*. https://jasp-stats.org/

Jürgens, R., Grass, A., Drolet, M., & Fischer, J. (2015). Effect of Acting Experience on Emotion Expression and Recognition in Voice: Non-Actors Provide Better Stimuli than Expected. *Journal of Nonverbal Behavior*, *39*(3), 195–214. https://doi.org/10.1007/s10919-015-0209-5

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814. https://doi.org/10.1037/0033-2909.129.5.770

Kachlicka, M., & Tierney, A. (2024). Voice actors show enhanced neural tracking of pitch, prosody perception, and music perception. *Cortex*, *178*, 213–222. https://doi.org/10.1016/j.cortex.2024.06.016

Kawahara, H., Morise, M., & Skuk, V. G. (2013). Temporally variable multi-aspect N-way morphing based on interference-free speech representations. *IEEE International Conference on Acoustics, Speech and Signal Processing*.

Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., & Banno, H. (2008). TANDEM-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. *IEEE International Conference on Acoustics, Speech and Signal Processing*.

Kawahara, H., & Skuk, V. G. (2018). Voice Morphing. In S. Frühholz, P. Belin, & K. R. Scherer (Eds.), *The Oxford Handbook of Voice Perception* (pp. 684–706). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198743187.013.31

Kleber, B [B.], Veit, R., Birbaumer, N., Gruzelier, J., & Lotze, M. (2010). The brain of opera singers: Experience-dependent changes in functional activation. *Cerebral Cortex*, *20*(5), 1144–1152. https://doi.org/10.1093/cercor/bhp177

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*(8), 599–605. https://doi.org/10.1038/nrn2882

Krishnan, S., Lima, C. F., Evans, S., Chen, S., Guldner, S., Yeff, H., Manly, T., & Scott, S. K. (2018). Beatboxers and Guitarists Engage Sensorimotor Regions Selectively When Listening to the Instruments They can Play. *Cerebral Cortex*, *28*(11), 4063–4079. https://doi.org/10.1093/cercor/bhy208

Laukka, P., Elfenbein, H. A., Thingujam, N. S., Rockstuhl, T., Iraki, F. K., Chui, W., & Althoff, J. (2016). The expression and recognition of emotions in the voice across five nations: A lens model analysis based on acoustic features. *Journal of Personality and Social Psychology*, *111*(5), 686–705. https://doi.org/10.1037/pspi0000066

Law, L. N. C., & Zentner, M. (2012). Assessing musical abilities objectively: Construction and validation of the profile of music perception skills. *PLoS One*, *7*(12), e52508. https://doi.org/10.1371/journal.pone.0052508

Lehnen, J. M., Schweinberger, S. R., & Nussbaum, C. (2025). Vocal Emotion Perception and Musicality-Insights from EEG Decoding. *Sensors (Basel, Switzerland)*, *25*(6). https://doi.org/10.3390/s25061669

Lima, C. F., Brancatisano, O., Fancourt, A., Müllensiefen, D., Scott, S. K., Warren, J. D., & Stewart, L. (2016). Impaired socio-emotional processing in a developmental music disorder. *Scientific Reports*, *6*, 34911. https://doi.org/10.1038/srep34911

Lima, C. F., & Castro, S. L. (2011). Speaking to the trained ear: Musical expertise enhances the recognition of emotions in speech prosody. *Emotion*, *11*(5), 1021–1031. https://doi.org/10.1037/a0024521

Lotze, M., Scheler, G., Tan, H.-R. M., Braun, C., & Birbaumer, N. (2003). The musician's brain: Functional imaging of amateurs and professionals during performance and imagery. *Neuroimage*, *20*(3), 1817–1829. https://doi.org/10.1016/j.neuroimage.2003.07.018

Loveday, C., Musgrave, G., & Gross, S.-A. (2023). Predicting anxiety, depression, and wellbeing in professional and nonprofessional musicians. *Psychology of Music*, *51*(2), 508–522. https://doi.org/10.1177/03057356221096506

Ly, A., Verhagen, J., & Wagenmakers, E.-J. (2016). Harold Jeffreys's default Bayes factor hypothesis tests: Explanation, extension, and application in psychology. *Journal of Mathematical Psychology*, *72*, 19–32. https://doi.org/10.1016/j.jmp.2015.06.004

Maghiar, M. J., Lawrence, B. J., Mulders, W. H., Moyle, T. C., Livings, I., & Jayakody, D. M. P. (2023). Hearing loss and mental health issues in amateur and professional musicians. *Psychology of Music*, *51*(6), 1584–1597. https://doi.org/10.1177/03057356231155970

Martins, I., Lima, C. F., & Pinheiro, A. P. (2022). Enhanced salience of musical sounds in singers and instrumentalists. *Cognitive, Affective, & Behavioral Neuroscience.* Advance online publication. https://doi.org/10.3758/s13415-022-01007-x

Martins, M., Pinheiro, A. P., & Lima, C. F. (2021). Does Music Training Improve Emotion Recognition Abilities? A Critical Review. *Emotion Review*, *13*(3), 199–210. https://doi.org/10.1177/17540739211022035

Mithen, S., Morley, I., Wray, A., Tallerman, M., & Gamble, C. (2006). The Singing Neanderthals: The Origins of Music, Language, Mind and Body by Steven Mithen. London: Weidenfeld & Nicholson, 2005. ISBN 0-297-64317-7 hardback £20 & US$25.2; ix+374 pp. *Cambridge Archaeological Journal*, *16*(1), 97–112. https://doi.org/10.1017/S0959774306000060

Moisseinen, N., Ahveninen, L., Martínez-Molina, N., Sairanen, V., Melkas, S., Kleber, B [Boris], Sihvonen, A. J., & Särkämö, T. (2024). Choir singing is associated with enhanced structural connectivity across the adult lifespan. *Human Brain Mapping*, *45*(7), e26705. https://doi.org/10.1002/hbm.26705

Morrison, S. J., & Demorest, S. M. (2009). Cultural constraints on music perception and cognition. *Progress in Brain Research*, *178*, 67–77. https://doi.org/10.1016/S0079-6123(09)17805-6

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS One*, *9*(2), e89642. https://doi.org/10.1371/journal.pone.0101091

Neves, L., Martins, M., Correia, A. I., Castro, S. L., Schellenberg, E. G., & Lima, C. F. (2025). Does music training improve emotion recognition and cognitive abilities? Longitudinal and correlational evidence from children. *Cognition*, *259*, 106102. https://doi.org/10.1016/j.cognition.2025.106102

Nikjeh, D. A., Lister, J. J., & Frisch, S. A. (2009). The relationship between pitch discrimination and vocal production: Comparison of vocal and instrumental musicians. *The Journal of the Acoustical Society of America*, *125*(1), 328–338. https://doi.org/10.1121/1.3021309

Nussbaum, C., Schirmer, A., & Schweinberger, S. R. (2022). Contributions of fundamental frequency and timbre to vocal emotion perception and their electrophysiological correlates. *Social Cognitive and Affective Neuroscience*, *17*(12), 1145–1154. https://doi.org/10.1093/scan/nsac033

Nussbaum, C., Schirmer, A., & Schweinberger, S. R. (2023). Electrophysiological Correlates of Vocal Emotional Processing in Musicians and Non-Musicians. *Brain Sciences*, *13*(11), 1563. https://doi.org/10.3390/brainsci13111563

Nussbaum, C., Schirmer, A., & Schweinberger, S. R. (2024). Musicality - Tuned to the melody of vocal emotions. *British Journal of Psychology*, *115*(2), 206–225. https://doi.org/10.1111/bjop.12684

Nussbaum, C., & Schweinberger, S. R. (2021). Links Between Musicality and Vocal Emotion Perception. *Emotion Review*, *13*(3), 211–224. https://doi.org/10.1177/17540739211022803

Oechslin, M. S., van de Ville, D., Lazeyras, F., Hauert, C.-A., & James, C. E. (2013). Degree of musical expertise modulates higher order brain functioning. *Cerebral Cortex*, *23*(9), 2213–2224. https://doi.org/10.1093/cercor/bhs206

Papadaki, E., Koustakas, T., Werner, A., Lindenberger, U., Kühn, S., & Wenger, E. (2023). Resting-state functional connectivity in an auditory network differs between aspiring professional and amateur musicians and correlates with performance. *Brain Structure & Function*, *228*(9), 2147–2163. https://doi.org/10.1007/s00429-023-02711-1

Patel, A. D. (2011). Why would Musical Training Benefit the Neural Encoding of Speech? The OPERA Hypothesis. *Front Psychol*, *2*, 142. https://doi.org/10.3389/fpsyg.2011.00142

R Core Team. (2025). *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/

Rakei, A., & Bhattacharya, J. (2024). Professional status matters: Differences in flow proneness between professional and amateur contemporary musicians. *Psychology of Aesthetics, Creativity, and the Arts.* Advance online publication. https://doi.org/10.1037/aca0000674

Rammstedt, B., Danner, D., Soto, C. J., & John, O. P. (2018). Validation of the short and extra-short forms of the Big Five Inventory-2 (BFI-2) and their German adaptations. *European Journal of Psychological Assessment.* Advance online publication. https://doi.org/10.1027/1015-5759/a000481

Rogenmoser, L., Kernbach, J., Schlaug, G., & Gaser, C. (2018). Keeping brains young with making music. *Brain Structure & Function*, *223*(1), 297–305. https://doi.org/10.1007/s00429-017-1491-2

Rosenfeld, J. P., & Olson, J. M. (2021). Bayesian Data Analysis: A Fresh Approach to Power Issues and Null Hypothesis Interpretation. *Applied Psychophysiology and Biofeedback*, *46*(2), 135–140. https://doi.org/10.1007/s10484-020-09502-y

Schellenberg, E. G., & Lima, C. F. (2024). Music Training and Nonmusical Abilities. *Annual Review of Psychology*, *75*, 87–128. https://doi.org/10.1146/annurev-psych-032323-051354

Scherer, K. R. (2018). Acoustic Patterning of Emotion Vocalizations. In S. Frühholz, P. Belin, & K. R. Scherer (Eds.), *The Oxford Handbook of Voice Perception* (pp. 60–92). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198743187.013.4

Schirmer, A., Croy, I., Liebal, K., & Schweinberger, S. R. (2025). Non-verbal effecting - Animal research sheds light on human emotion communication. *Biological Reviews of the Cambridge Philosophical Society*, *100*(1), 245–257. https://doi.org/10.1111/brv.13140

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096–1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*(1), 24–31. https://doi.org/10.1177/0098628316677643

Swaminathan, S., & Schellenberg, E. G. (2020). Musical ability, music training, and language ability in childhood. *J Exp Psychol Learn Mem Cogn*, *46*(12), 2340–2348. https://doi.org/10.1037/xlm0000798

Taylor, A., & Hallam, S. (2011). From leisure to work: amateur musicians taking up instrumental or vocal teaching as a second career. *Music Education Research*, *13*(3), 307–325. https://doi.org/10.1080/14613808.2011.603044

Thompson, W. F., Marin, M. M., & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences*, *109*(46), 19027–19032. https://doi.org/10.1073/pnas.1210344109

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion*, *4*(1), 46–64. https://doi.org/10.1037/1528-3542.4.1.46

Tragantzopoulou, P., & Giannouli, V. (2025). A Song for the Mind: A Literature Review on Singing and Cognitive Health in Aging Populations. *Brain Sciences*, *15*(3). https://doi.org/10.3390/brainsci15030227

Vigl, J., Talamini, F., Strauss, H., & Zentner, M. (2024). Prosodic discrimination skills mediate the association between musical aptitude and vocal emotion recognition ability. *Scientific Reports*, *14*(1), 16462. https://doi.org/10.1038/s41598-024-66889-y

Vincenzi, M., Correia, A. I., Vanzella, P., Pinheiro, A. P., Lima, C. F., & Schellenberg, E. G. (2022). Associations between music training and cognitive abilities: The special case of professional musicians. *Psychology of Aesthetics, Creativity, and the Arts.* Advance online publication. https://doi.org/10.1037/aca0000481

Waters, S., Kanber, E., Lavan, N., Belyk, M., Carey, D., Cartei, V., Lally, C., Miquel, M., & McGettigan, C. (2021). Singers show enhanced performance and neural representation of vocal imitation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *376*(1840), 20200399. https://doi.org/10.1098/rstb.2020.0399

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Zendel, B. R., & Alexander, E. J. (2020). Autodidacticism and Music: Do Self-Taught Musicians Exhibit the Same Auditory Processing Advantages as Formally Trained Musicians? *Frontiers in Neuroscience*, *14*, 752. https://doi.org/10.3389/fnins.2020.00752

Zentner, M., & Strauss, H. (2017). Assessing musical ability quickly and objectively: Development and validation of the Short-PROMS and the Mini-PROMS. *Annals of the New York Academy of Sciences*, *1400*(1), 33–45. https://doi.org/10.1111/nyas.13410

Vocal Emotion Recognition: A Comparison of Singers and Instrumentalists, Amateurs and Professionals

STAR Methods

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Participants for Part I and II

(reported in the manuscript in section “Part I - Participants”)

Data were collected in a pseudonymized format from June 2023 to January 2024. All participants were aged between 18 and 54 years and fluent German speakers. Participants provided informed consent before completing the experiment and received compensation in the form of 12.50 € or course credit upon completion. The experiment was in line with the ethical guidelines by the American Psychological Association (APA) and approved by the local ethics committee of the Friedrich Schiller University Jena (Reg.-Nr. FSV 19/045).

In total, we collected data from 94 amateur musicians that were divided into singers and instrumentalists. Recruitment criteria specified that participants had to be non-professional musicians (i.e., they held no music-related academic degree or worked professionally as a musician). Singers were required to be currently active in a choir or another singing group but should not play an instrument actively and regularly (i.e., they must not currently be instrumentalists in an orchestra or a band). Instrumentalists, conversely, were required to be currently active in an orchestra or a band, but they should not engage in singing activities actively and regularly (i.e., they must not currently be in a choir or another singing group).

Singers

We recorded data from 48 singers, of which three were excluded (N = 2 had > 5 % trials of omission, N = 1 had technical issues during stimulus playback). Thus, data from 45 singers were analyzed (22 female, 22 male, 1 diverse, aged 18 to 53 years [M = 27.02, SD = 8.2]). Mean onset age of musical training was 8 years (SD = 3.08, 5 - 20 years). Mean duration of musical training was 10 years (SD = 6.59, 0 – 25 years). Five participants reported that they never had any formal musical training. Two participants reported that they had occasional tinnitus, but without any subjective impairments in daily life.

Instrumentalists

Data from 46 instrumentalists were collected, of which three were excluded. One had technical issues during stimulus playback, one was also active in a choir, one held a master’s degree in music science and was therefore regrouped with the professional musicians (see Part III). Thus, data from 43 instrumentalists entered analysis (24 female, 18 male, 1 diverse, aged 18 to 54 years [M = 28.51, SD = 10.64]). Mean onset age of musical training was 7 years (SD = 2.27, 4 - 14 years). Mean duration of musical training was 14 years (SD = 10.00, 0 – 44 years). Four participants reported that they never had any formal musical training, thus were autodidacts.

For more details, please refer to the supplemental sample information in the document “Supplemental_Tables_and_Figures.pdf” in the associated OSF repository (https://osf.io/ascqx/): S1 for musical background, S2 for socioeconomic background.

Participants for Part III

(reported in the manuscript in section “Part III - Method”)

For this analysis, we collapsed all participants from Part I into the group of amateur musicians and compared it to the groups of professional musicians and non-musicians reported in Nussbaum et al. (2024). Note that we added one participant to the professional group, because he held a master’s degree in music (see Part I), so numbers slightly diverge from the original publication. All professional musicians reported to have a music-related academic degree or a non-academic music qualification. Non-musicians played no instrument or engaged in any other musical activities. For a more detailed description, please refer to Nussbaum et al. (2024).

In total, we analyzed data from 40 professional musicians (20 male, 20 female, aged 20 to 42 years [M = 29.6; SD = 5.58]), 38 non-musicians (18 male, 20 female, aged 19 to 48 years [M = 30.5; SD = 6.54]) and 88 amateurs (40 male, 46 female, 2 diverse, aged 18 to 54 years [M = 27.8; SD = 9.44]).

METHOD DETAILS

Stimulus material

(reported in the manuscript in section “Part I – Stimulus material”)

As stimulus material, we used parameter-specific voice morphs that express emotional information either through the fundamental frequency contour only (F0), through timbre only (Tbr) or through a combination of both (Full).

For voice morphing, we selected original audio recordings from a database of vocal actor portrayals, comprised of pseudowords (/molen/, /loman/, /belam/) uttered by eight speakers (four male, four female) with expressions of happiness, pleasure, fear, and sadness. We specifically opted for two positive and two negative emotions of different intensities, to balance both valence and arousal. A prior validation study with 20 raters confirmed that the two positive and two negative emotions had different degrees of emotional intensity (happiness > pleasure, $t(19)=9.57$, $p < 0.001$ and fear > sadness, $t(19)=6.58$, $p < 0.001$).

To synthesize the parameter-specific emotional voice morphs, we created morphing trajectories between each emotion and an emotional average of the same speaker and pseudoword, using the Tandem-STRAIGHT software (Kawahara et al., 2013; Kawahara et al., 2008). The averages had been created previously by blending all emotions together and were thus assumed to be uninformative and unbiased with respect to the four emotions of interest. After substantial preprocessing (e.g. manual mapping of time- and frequency anchors in each stimulus), Tandem-STRAIGHT enables voice morphing via weighted interpolation of five independent parameters: (1) F0-contour, (2) timing, (3) spectrum-level, (4) aperiodicity, and (5) spectral frequency; the latter three are summarized as timbre.

We created three types of morphed stimuli. Full-Morphs were stimuli with all parameters taken from the emotional version (corresponding to 100% from the emotion and 0% from average), except for the timing parameter, which was taken from the average (corresponding to 0% emotion and 100% average). F0-Morphs were stimuli with the F0-contour taken from the emotion, but timbre and timing were taken from the average. Timbre-Morphs were stimuli with all timbre parameters taken from the emotion, but F0 and timing from the average. Note that the timing was kept constant in all conditions to allow a pure comparison of F0 vs. timbre. Furthermore, we kept all average stimuli as a further ambiguous reference category. In total, this resulted in 8 (speakers) x 3 (pseudowords) x 4 (emotions) x 3 (morphing conditions) + 24 average (8 speakers x 3 pseudowords) = 312 stimuli (duration M = 780 ms, range 620 to 967 ms, SD = 98 ms). Using PRAAT (Boersma, 2018), we normalized all stimuli to a root-mean-square of 70 dB SPL.

For a more detailed description of the stimulus creation, see Nussbaum et al. (2024) and Kawahara and Skuk (2018).

For a summary of acoustic characteristics, see Tables S3 and S4 in the document "Supplemental_Tables_and_Figures.pdf" in the associated OSF repository (https://osf.io/ascqx/)

Design

(reported in the manuscript in section "Part I – Design")

Data were collected online via PsyToolkit (Stoet, 2010, 2017), but after completion of the study all participants met with the experimenter for a short personal debriefing. This was done to increase commitment and conscientiousness for the experiment.

Participants were required to ensure a quiet environment for the duration of the study and use a computer with a physical keyboard and headphones. As browser, we recommended Google Chrome, and excluded Safari for technical reasons. Prior to the listening tasks, participants could adjust their sound settings to a comfortable sound pressure level.

First, participants entered demographic information, including age, sex, native language, profession, and potential hearing impairments such as tinnitus. They then completed an emotion classification experiment, a test on music perception and several questionnaires on musicality, personality and socioeconomic background. Mean duration of the whole online study was about 75 minutes.

Emotion classification experiment

In the experiment, participants classified vocal emotions as happiness, pleasure, fear, or sadness. Each trial started with a green fixation cross presented for 500 ms. Then the sound was played while a loudspeaker symbol was shown on the screen. Subsequently, a response screen showed the emotion labels and participants could enter their response within a 5000 ms time window starting from voice offset. Responses were entered with the left and right index and middle fingers, with random mapping of response keys to emotion categories for each participant, out of four possible key mappings (see Tables S5 and S6 in the document "Supplemental_Tables_and_Figures.pdf" in the associated OSF repository). In case of a response omission, the final trial slide (500 ms) prompted participants to respond faster; otherwise, the screen turned black. Then the next trial started.

The 312 stimuli were presented in randomized order in six blocks of 52 trials each, with self-paced breaks in between. Beforehand, participants completed eight practice trials with different stimuli. The experiment was about 25 minutes long. Unfortunately, due to a software error, randomization was sampled with replacements, so that some stimuli were drawn repeatedly and others were omitted. This was in contrast to our previous study, where randomization was sampled without replacement so that each stimulus was drawn exactly once.

Profile of Music Perception Skills (PROMS)

To measure music perception skills, we used a modular version of the Profile of Music Perception Skills (Law & Zentner, 2012; Zentner & Strauss, 2017), comprised of the four subtests „Melody“, „Pitch“, „Timbre”, and „Rhythm“, which we considered most informative for the present research question. Participants completed 18 items per subtest, always preceded by one practice trial. Each trial, participants heard a reference stimulus twice followed by a target stimulus. Then, they indicated whether reference and target were the same or different via a 5-point Likert scale with the labels “definitely same”, “maybe same”, “don’t know”, “maybe different”, and “definitely different”. The duration of the PROMS was about 20 minutes.

Questionnaires

After the PROMS, participants completed several questionnaires: the German Version of the Autism Quotient Questionnaire, AQ, (Baron-Cohen et al., 2001; Freitag et al., 2007), a 30-item Personality Inventory measuring the Big-Five domains (Rammstedt et al., 2018), the Goldsmiths Musical Sophistication Index, Gold-MSI, (Müllensiefen et al., 2014) to assess the participants’ degree of self-reported musical skills, additional questions concerning music experience and musical engagement, their socioeconomic background, and the 20-item version of the Positive-Affect-Negative-Affect-Scale, PANAS (Breyer & Bluemke, 2016; Watson et al., 1988).

QUANTIFICATION AND STATISTICAL ANALYSIS

(reported in the manuscript in section “Part I – Data analysis”)

In line with our preregistered plan, we collapsed data across speakers and pseudowords for analysis. Further, data on emotional averages were excluded because they were not relevant for our hypotheses. Response omissions (~1 %) were treated as errors and participants with more than 5% of such omissions excluded from data analysis. Analyses of Variance (ANOVAs) and correlational analyses were performed using R Version 4.5.0 (R Core Team, 2025). Post-hoc tests were Benjamini-Hochberg corrected where appropriate (Benjamini & Hochberg, 1995). A p-value < 0.05 was considered significant.

We complemented these classical frequentist analyses with a Bayesian approach, which – in contrast to null hypothesis significance testing - allows a quantification of evidence for null findings (Rosenfeld & Olson, 2021). These analyses were conducted in JASP Version 0.19.3 (JASP Team, 2025) using default priors, which have been considered appropriate for testing null hypotheses based on similar sample sizes (Ly et al., 2016; Neves et al., 2025). Further, we ensured that our Bayesian inference did not depend critically on the choice of priors by running robustness checks (see data analysis files “Analysis_JASP_Bayesian.zip” on OSF). We report the Bayes factor (BF10) as an indicator for the likelihood of the null and alternative hypothesis given the observed data. BF10 > 1 indicate larger evidence for the alternative hypothesis, BF10 < 1 indicate larger evidence for the null hypothesis. For example, a BF10 = 3 means that the alternative hypothesis is three times more likely than the null hypothesis, whereas the reciprocal BF10 = .33 means that the null hypothesis is three times more likely than the alternative. Following the guidelines by Jarosz and Wiley (2014), we consider values of BF10 = 1-3 (.1-.33) as anecdotal, BF10 = 3-10 (.33-.10) as moderate, BF10 = 10-30 (.10-

.03) as strong, BF10 = 30-100 (.03-.01) as very strong and BF10 > 100 (< .01) as decisive evidence for the alternative hypothesis and the reciprocal values in parentheses as respective evidence for the null hypothesis.

In alignment with the approach by Nussbaum et al. (2024), we first recoded responses in the PROMS from 1 to 0 in 0.25 steps starting with the "definitely" correct option down to the "definitely" incorrect option (thus, "don't know" was always coded with 0.5). For the final measure, we then subtracted 0.5, so that a positive score indicates that participants were more correct/confident, a negative score indicates more incorrect/uncertain ratings, and a score of zero indicates responses at chance level. For statistical analyses, we used the averaged performance across trials for each subtest.

ADDITIONAL RESOURCES

(reported in the manuscript in section "Part I – Transparency and openness")

We specified how we determined our sample size, all data exclusions, all manipulations, and all measures in the associated preregistration (https://doi.org/10.17605/OSF.IO/76PV5). Preprocessed data, analysis scripts and supplemental materials can be found in the associated OSF repository (https://osf.io/ascqx/).

KEY RESOURCES TABLE

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
a zip-Folder with Stimulus examples: "stimulus_examples.zip"	Associated OSF repository provided by the authors	https://osf.io/ascqx/
a zip-Folder containing the frequentist analysis + preprocessed data in R: "Analysis_R_Frequentist.zip"	Associated OSF repository provided by the authors	https://osf.io/ascqx/
a zip-Folder containing the Bayesian analysis in JASP: "Analysis_JASP_Bayesian.zip"	Associated OSF repository provided by the authors	https://osf.io/ascqx/
a PDF with supplemental Figures and Tables: "Supplemental_Tables_and_Figures.pdf"	Associated OSF repository provided by the authors	https://osf.io/ascqx/
Software and algorithms		
R Version 4.5.0 (used for statistical analysis)	R Core Team (2025)	https://cran.r-project.org/
JASP Version 0.19.3 (used for statistical analysis)	JASP Team (2025)	https://jasp-stats.org/

References

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J.-C., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Boersma, P. (2018). Praat: Doing phonetics by computer [Computer program]: Version 6.0.46, retrieved January 2020 from http://www.praat.org/. *Http://www. Praat. Org*.

Breyer, B., & Bluemke, M. (2016). *Deutsche Version der Positive and Negative Affect Schedule PANAS (GESIS Panel).* https://doi.org/10.6102/zis242

Freitag, C. M., Retz-Junginger, P., Retz, W., Seitz, C., Palmason, H., Meyer, J., Rösler, M., & Gontard, A. von (2007). Evaluation der deutschen Version des Autismus-Spektrum-Quotienten (AQ) - die Kurzversion AQ-k. *Zeitschrift Für Klinische Psychologie Und Psychotherapie*, *36*(4), 280–289. https://doi.org/10.1026/1616-3443.36.4.280

Jarosz, A. F., & Wiley, J. (2014). What Are the Odds? A Practical Guide to Computing and Reporting Bayes Factors. *The Journal of Problem Solving*, *7*(1). https://doi.org/10.7771/1932-6246.1167

JASP Team. (2025). *JASP (Version 0.19.3)[Computer software]*. https://jasp-stats.org/

Kawahara, H., Morise, M., & Skuk, V. G. (2013). Temporally variable multi-aspect N-way morphing based on interference-free speech representations. *IEEE International Conference on Acoustics, Speech and Signal Processing*.

Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., & Banno, H. (2008). TANDEM-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. *IEEE International Conference on Acoustics, Speech and Signal Processing*.

Kawahara, H., & Skuk, V. G. (2018). Voice Morphing. In S. Frühholz, P. Belin, S. Frühholz, P. Belin, & K. R. Scherer (Eds.), *The Oxford Handbook of Voice Perception* (pp. 684–706). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198743187.013.31

Law, L. N. C., & Zentner, M. (2012). Assessing musical abilities objectively: Construction and validation of the profile of music perception skills. *PLoS One*, *7*(12), e52508. https://doi.org/10.1371/journal.pone.0052508

Ly, A., Verhagen, J., & Wagenmakers, E.-J. (2016). Harold Jeffreys's default Bayes factor hypothesis tests: Explanation, extension, and application in psychology. *Journal of Mathematical Psychology*, *72*, 19–32. https://doi.org/10.1016/j.jmp.2015.06.004

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS One*, *9*(2), e89642. https://doi.org/10.1371/journal.pone.0101091

Neves, L., Martins, M., Correia, A. I., Castro, S. L., Schellenberg, E. G., & Lima, C. F. (2025). Does music training improve emotion recognition and cognitive abilities? Longitudinal and correlational evidence from children. *Cognition*, *259*, 106102. https://doi.org/10.1016/j.cognition.2025.106102

Nussbaum, C., Schirmer, A., & Schweinberger, S. R. (2024). Musicality - Tuned to the melody of vocal emotions. *British Journal of Psychology*, *115*(2), 206–225. https://doi.org/10.1111/bjop.12684

R Core Team. (2025). *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/

Rammstedt, B., Danner, D., Soto, C. J., & John, O. P. (2018). Validation of the short and extra-short forms of the Big Five Inventory-2 (BFI-2) and their German adaptations. *European Journal of Psychological Assessment.* Advance online publication. https://doi.org/10.1027/1015-5759/a000481

Rosenfeld, J. P., & Olson, J. M. (2021). Bayesian Data Analysis: A Fresh Approach to Power Issues and Null Hypothesis Interpretation. *Applied Psychophysiology and Biofeedback*, *46*(2), 135–140. https://doi.org/10.1007/s10484-020-09502-y

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096–1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*(1), 24–31. https://doi.org/10.1177/0098628316677643

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Zentner, M., & Strauss, H. (2017). Assessing musical ability quickly and objectively: Development and validation of the Short-PROMS and the Mini-PROMS. *Annals of the New York Academy of Sciences*, *1400*(1), 33–45. https://doi.org/10.1111/nyas.13410

CELL PRESS DECLARATION OF INTERESTS POLICY

Transparency is essential for a reader's trust in the scientific process and for the credibility of published articles. At Cell Press, we feel that disclosure of competing interests is a critical aspect of transparency. Therefore, we require a "declaration of interests" section in which all authors disclose any financial or other interests related to the submitted work that (1) could affect or have the perception of affecting the author's objectivity or (2) could influence or have the perception of influencing the content of the article.

<u>What types of articles does this apply to?</u>
We require that you disclose competing interests for all submitted content by completing and submitting the form below. We also require that you include a "declaration of interests" section in the text of all articles even if there are no interests to declare.

<u>What should I disclose?</u>
We require that you and all authors disclose any personal financial interests (e.g., stocks or shares in companies with interests related to the submitted work or consulting fees from companies that could have interests related to the work), professional affiliations, advisory positions, board memberships (including membership on a journal's advisory board when publishing in that journal), or patent applications and/or registrations that are related to the subject matter of the contribution. As a guideline, you need to declare an interest for (1) any affiliation associated with a payment or financial benefit exceeding $10,000 p.a. or 5% ownership of a company or (2) research funding by a company with related interests. You do not need to disclose diversified mutual funds, 401ks, or investment trusts.

Authors should also disclose relevant financial interests of immediate family members. Cell Press uses the Public Health Service definition of "immediate family member," which includes spouse and dependent children.

<u>Where do I declare competing interests?</u>
Competing interests should be disclosed on this form as well as in a "declaration of interests" section in the manuscript. This section should include financial or other competing interests as well as affiliations that are not included in the author list. Examples of "declaration of interests" language include:

> "AUTHOR is an employee and shareholder of COMPANY."
> "AUTHOR is a founder of COMPANY and a member of its scientific advisory board."

NOTE: Primary affiliations should be included with the author list and do not need to be included in the "declaration of interests" section. Funding sources should be included in the "acknowledgments" section and also do not need to be included in the "declaration of interests" section. (A small number of front-matter article types do not include an "acknowledgments" section. For these articles, reporting of funding sources is not required.)

<u>What if there are no competing interests to declare?</u>
If you have no competing interests to declare, please note that in the "declaration of interests" section with the following wording:

> "The authors declare no competing interests."

April 12, 2024

CELL PRESS DECLARATION OF INTERESTS FORM

If submitting materials via Editorial Manager, please complete this form and upload with your initial submission. Otherwise, please email as an attachment to the editor handling your manuscript.

Please complete each section of the form and insert any necessary "declaration of interests" statement in the text box at the end of the form. A matching statement should be included in a "declaration of interests" section in the manuscript.

Institutional affiliations

We require that you list the current institutional affiliations of all authors, including academic, corporate, and industrial, on the title page of the manuscript. ***Please select one of the following:***

☒ All affiliations are listed on the title page of the manuscript.

☐ I or other authors have additional affiliations that we have noted in the "declaration of interests" section of the manuscript and on this form below.

Funding sources

We require that you disclose all funding sources for the research described in this work. ***Please confirm the following:***

☒ All funding sources for this study are listed in the "acknowledgments" section of the manuscript.*

*A small number of front-matter article types do not include an "acknowledgments" section. For these, reporting funding sources is not required.

Competing financial interests

We require that authors disclose any financial interests and any such interests of immediate family members, including financial holdings, professional affiliations, advisory positions, board memberships, receipt of consulting fees, etc., that:

(1) could affect or have the perception of affecting the author's objectivity, *or*
(2) could influence or have the perception of influencing the content of the article.

Please select one of the following:

☒ We, the authors and our immediate family members, have no financial interests to declare.

☐ We, the authors, have noted any financial interests in the "declaration of interests" section of the manuscript and on this form below, and we have noted interests of our immediate family members.

Advisory/management and consulting positions
We require that authors disclose any position, be it a member of a board or advisory committee or a paid consultant, that they have been involved with that is related to this study. We also require that members of our journal advisory boards disclose their position when publishing in that journal. ***Please select one of the following:***

☒ We, the authors and our immediate family members, have no positions to declare and are not members of the journal's advisory board.

☐ The authors and/or their immediate family members have management/advisory or consulting relationships noted in the "declaration of interests" section of the manuscript and on this form below.

Patents
We require that you disclose any patent applications and/or registrations related to this work by any of the authors or their institutions. ***Please select one of the following:***

☒ We, the authors and our immediate family members, have no related patent applications or registrations to declare.

☐ We, the authors, have a patent application and/or registration related to this work, which is noted in the "declaration of interests" section of the manuscript and on this form below, and we have noted the patents of immediate family members.

Please insert any "declaration of interests" statements in this space. This exact text should also be included in the "declaration of interests" section of the manuscript. If no authors have a competing interest, please insert the text, "The authors declare no competing interests."

The authors declare no competing interests.

☒ **On behalf of all authors, I declare that I have disclosed all competing interests related to this work. If any exist, they have been included in the "declaration of interests" section of the manuscript.**